

Alma Mater Studiorum Università di Bologna

DIPARTIMENTO DI INTERPRETAZIONE E TRADUZIONE

Corso di Laurea magistrale Specialized Translation (classe LM - 94)

TESI DI LAUREA

**in RESEARCH METHODS FOR TRANSLATION, INTERPRETING AND
INTERCULTURAL STUDIES**

**Stereotipi di genere e traduzione automatica
dall'inglese all'italiano: uno studio di caso sul
femminile nelle professioni**

CANDIDATA:

Alessandra Luccioli

RELATRICE:

Silvia Bernardini

CORRELATRICE:

Raffaella Baccolini

Anno Accademico 2018/2019

Terzo Appello

Sommario

Abstract.....	1
Resumen.....	2
Introduzione	3
Capitolo I: Traduzione automatica e questioni di genere	5
1.1 Introduzione	5
1.2. Traduzione automatica: definizione e breve panoramica storica.....	5
1.3 Architetture dei sistemi di TA	8
1.3.1 Premessa	8
1.3.2 Architetture <i>rule-based</i>	9
1.3.3 Architetture statistiche	9
1.3.4 Architetture neurali	10
1.3.4.1 Reti Neurali Ricorrenti (RNN)	12
1.3.4.2 Reti Neurali Convoluzionali (CNN)	13
1.3.5 Intelligenza artificiale e società	13
1.3.5.1 Nuove applicazioni per la traduzione automatica.....	14
1.3.6 <i>Mind the gap</i> : il gap di genere nell’elaborazione del linguaggio naturale	16
1.3.6.1 Alcuni esempi di studi sugli stereotipi di genere nel NLP	17
1.3.6.2 Stereotipi di genere e traduzione automatica	19
1.3.6.3 Stereotipi di genere e traduzione automatica: il focus sull’italiano	22
Capitolo II: genere e lingua.....	24
2.1. Introduzione	24
2.2 Stereotipi e pregiudizi	24
2.3 Lingua e genere: il caso dei femminili professionali	27
2.4 La questione dei femminili professionali in Italia	29
2.5 La questione dei femminili professionali in Francia	32

2.6 La questione dei femminili professionali in Spagna.....	35
2.6 La questione nella lingua inglese.....	37
2.7 Conclusione e nuovi spunti.....	38
Capitolo III: Metodo.....	40
3.1 Studi precedenti	40
3.2 Struttura frasale.....	44
3.3 Selezione delle professioni	44
3.4 Confronto con dati italiani	49
3.5 Analisi <i>corpus-based</i>	51
Capitolo IV: Analisi dello studio	53
4.1 Presentazione dello studio	53
4.2 Analisi dei dati	53
4.2.1 Antecedente	53
4.2.2 Professione.....	53
4.2.3 Subset professioni maschili	54
4.2.4 Subset professioni femminili	56
4.2.5 Analisi sintagma nominale.....	58
4.2.6 Confronto tra i due set e numero totale di errori.....	60
4.3 Approfondimento sulle traduzioni dei termini chef/cuoco/cuoca.....	61
4.3.1 Analisi dei collocati di chef	62
4.3.2 Analisi dei collocati di cuoco.....	65
4.3.3 Analisi dei collocati di cuoca.....	68
4.4 Approfondimento sulle traduzioni dei termini contabile/commercialista	71
4.5 Approfondimento sull'output fornito da Google Translate	71
4.6 Discussione	74
Conclusioni	76
Bibliografia	78

Sitografia.....83

Allegati.....87

Abstract

New advances in machine translation systems have substantially improved the quality of translation output, yet many issues still need to be addressed: one of these is the presence of gender bias in machine translation systems. The aim of this dissertation is to investigate the relation between gender bias and machine translation through a case study employing sentences containing nouns referred to professions. The first chapter provides a theoretical overview of machine translation and then outlines the issue of gender bias in machine translation. The second chapter focuses on the relationship between gender and language, starting with the definition of the notions of stereotype and prejudice, and then discussing the issue of female forms of nouns referring to occupations. The linguistic aspects are outlined, and a brief overview is provided of the initiatives that have been proposed in the Italian and international context to promote the use of the female gender in language and to avoid sexist and non-inclusive language use. In the third chapter, after explaining in detail the methodology employed in previous studies, the sentence pattern used in the case study is presented. Statistical data on the number of women for each occupation are elaborated, the selected professions for the case study are provided, and organized in tables. Furthermore, the tools used in the corpus-based analysis carried out in chapter four are outlined. This chapter is devoted to the analysis of the outputs of DeepL and Google Translate, two machine translation systems, from English into Italian. In addition to the detailed analysis of all the linguistic aspects of the sentence pattern, explanatory tables and graphs are provided, as well as an in-depth analysis of the translations of some particularly relevant professions. Finally, the results of the analysis are discussed in depth, and future research perspectives in this field are proposed.

Resumen

El objetivo de esta tesis es investigar la relación entre los estereotipos de género y la traducción automática, proponiendo un estudio de caso consistente en frases que contienen sustantivos que se refieren a profesiones traducidas automáticamente del inglés al italiano. En el capítulo I ofreceré una visión general teórica de la traducción automática, y presentaré el tema de los estereotipos de género en la traducción automática. En el capítulo II me concentraré en la relación entre el género y el lenguaje a partir de las definiciones de “estereotipo” y “prejuicio”; sucesivamente, profundizaré en la cuestión de los femeninos de las profesiones. Al respecto, describiré los aspectos lingüísticos correspondientes y ofreceré una breve reseña de las iniciativas emprendidas en el contexto italiano e internacional para promover el uso del género femenino en el idioma y evitar su utilización de manera sexista y no inclusiva. En el capítulo III, tras explicar en detalle la metodología utilizada en los estudios anteriores, presentaré el diseño de las frases pensado para llevar a cabo el estudio de casos. A partir de la elaboración de los datos estadísticos sobre el número de mujeres para cada ocupación, presentaré las profesiones empleadas en el estudio. En el capítulo IV presentaré los instrumentos utilizados en el análisis basado en corpus, en el que se analizan los resultados proporcionados por dos sistemas de traducción automática, DeepL y Google Translate, entre inglés e italiano. El análisis detallado de todos los aspectos de la estructura de las frases va acompañado de tablas y gráficos explicativos, así como de un análisis de las traducciones de algunas profesiones especialmente relevantes. A continuación, discutiré los resultados del análisis de manera detallada y propondré posibles líneas de estudio y perspectivas de investigación futuras en este ámbito.

Introduzione

I recenti sviluppi nel campo della traduzione automatica, grazie all'impiego di reti neurali, hanno reso i sistemi sempre più accurati e in grado di decodificare frasi lunghe e strutture sintattiche sempre più complesse. Allo stesso tempo, sempre più studi si stanno concentrando sulle questioni legate all'equità e agli stereotipi presenti nell'elaborazione del linguaggio naturale, processo nel quale rientra anche l'impiego dei sistemi di traduzione automatica. L'idea alla base di questa tesi è nata dalla lettura di un articolo dal sito *Slator*¹ (specializzato in ricerca e traduzione) pubblicato il 22 gennaio 2019 nel quale venivano riportati gli sforzi del colosso dell'informatica Google per rendere meno sessista il suo sistema di traduzione automatica, Google Translate. L'azienda, in un articolo pubblicato nel dicembre 2018, annunciava che oltre ai suoi sforzi per “promuovere l'equità e ridurre gli stereotipi nell'apprendimento automatico” (Kuczmarski, 2018: online), si sarebbe impegnata a ridurre gli stereotipi di genere nel suo sistema di traduzione automatica, Google Translate, fornendo traduzioni femminili e maschili per alcune parole o brevi frasi in alcune combinazioni linguistiche. Questo articolo ha catturato la mia attenzione perché fino a quel momento non mi ero mai soffermata a pensare che i sistemi di traduzione automatica, e in generale le applicazioni dell'intelligenza artificiale alla lingua, potessero presentare degli stereotipi o un linguaggio discriminatorio. Questo interesse mi ha portato ad approfondire l'argomento anche grazie al corso di *Research Methods*, il cui esame è stato un primo “incubatore” di questo elaborato. Lo scopo di questa tesi è quindi quello di investigare la presenza di stereotipi di genere, e dunque l'eventuale mancanza di attenzione rispetto ad un uso non sessista della lingua, nei sistemi di traduzione automatica di uso comune. Il presente elaborato espone i due argomenti alla base della ricerca: la traduzione automatica e il rapporto tra stereotipi di genere e lingua, con una particolare attenzione alla questione dei femminili professionali. Successivamente viene delineato il metodo utilizzato nello studio e l'analisi dei risultati ottenuti e infine esposte le possibili prospettive future in quest'ambito di ricerca.

In particolare, il capitolo I offre una panoramica teorica sulla traduzione automatica, dalle sue origini fino ai sistemi attualmente utilizzati. Sono poi esposte le varie architetture dei sistemi di traduzione automatica, con una particolare attenzione alle

¹ <https://slator.com/technology/he-said-she-said-addressing-gender-in-neural-machine-translation/>

architetture neurali, attualmente impiegate nella maggior parte dei sistemi disponibili online. Successivamente si espone il rapporto tra gli stereotipi di genere e l'elaborazione del linguaggio naturale, soffermandosi in dettaglio sugli studi che affrontano la questione dal punto di vista della traduzione automatica.

Il capitolo II si concentra invece sul rapporto tra genere e lingua, partendo dalla definizione di stereotipo e pregiudizio per passare poi alla questione dei femminili professionali. Vengono delineati gli aspetti linguistici per poi fornire una breve panoramica delle iniziative che sono state proposte nel contesto italiano per promuovere l'uso del genere femminile nella lingua ed evitarle di usarla in maniera sessista e non inclusiva. Si passa poi a presentare le azioni intraprese in questo ambito da paesi come Francia e Spagna, sia per la vicinanza linguistica tra la lingua italiana e le altre due lingue romanze sia per confrontare diversi approcci alla questione. Infine si presentano le linee guida e le riflessioni che sono state portate avanti per la lingua inglese, che, pur non presentando marcatura di genere, non è esente da usi sessisti o non paritari della lingua.

Il capitolo III espone la metodologia impiegata nello studio proposto. Vengono prima illustrati gli studi precedenti in materia e le motivazioni che hanno portato alla scelta di una struttura frasale diversa da quelle presentate in altri studi. Si illustrano poi nel dettaglio i criteri per la selezione delle professioni sulla base dei dati elaborati dallo *US Bureau of Labor Force Statistics*, che mostrano il numero di donne per ogni occupazione. Questi dati vengono quindi confrontati con quelli italiani in modo da fornire una rappresentazione statistica valida per entrambe le lingue dello studio. Infine si propone una descrizione degli strumenti utilizzati nell'analisi basata su corpora condotta nel successivo capitolo.

Il capitolo IV presenta invece l'analisi degli output forniti dai due sistemi di traduzione automatica impiegati nello studio, DeepL e Google Translate, nella combinazione linguistica inglese-italiano. Si valuteranno in maniera dettagliata tutti gli aspetti della frase, corredando l'analisi con tabelle e grafici esplicativi. Inoltre si presentano degli approfondimenti sulle traduzioni di alcune professioni che hanno una particolare rilevanza e vengono inoltre approfondite le traduzioni dei termini relativi alle occupazioni forniti da Google Translate. I risultati dell'analisi sono infine discussi in maniera approfondita e globale.

Capitolo I: Traduzione automatica e questioni di genere

1. Introduzione

La traduzione automatica è una delle tecnologie che più è progredita negli ultimi anni, tra quelle a supporto del processo traduttivo. In questo capitolo ne verrà ripercorsa sommariamente la storia, dalla nascita agli sviluppi più recenti. Verranno poi delineate le principali architetture utilizzate nella traduzione automatica, rivolgendo una particolare attenzione ai sistemi neurali, lo standard utilizzato oggi, che più ha contribuito a rendere questa tecnologia una possibile alternativa alla traduzione manuale in diversi ambiti e generi. Dato il campo di applicazione di questo lavoro, ovvero lo studio degli stereotipi di genere nella traduzione automatica tra l'inglese e l'italiano, verranno esposti alcuni progetti di ricerca che utilizzano la traduzione automatica per scopi umanitari o per finalità utili alla società, nonché alcuni studi che si concentrano sugli stereotipi di genere nei campi di applicazione dell'elaborazione del linguaggio naturale, e in particolare nella traduzione automatica.

1.2. Traduzione automatica: definizione e breve panoramica storica

Il termine “traduzione automatica” (TA) o, in inglese, “*machine translation*” (MT) viene utilizzato per riferirsi a “sistemi computerizzati responsabili della produzione di traduzioni da una lingua naturale all'altra, con o senza l'ausilio di persone” (Hutchins e Somers, 1992:3, traduzione mia²).

Questa definizione, nonostante sia stata elaborata nei primi anni '90, espone ancora chiaramente la finalità di questa tecnologia, quella di tradurre tra lingue naturali, ovvero utilizzate per la comunicazione tra esseri umani, in contrapposizione al linguaggio informatico di programmazione (Gaspari 2011:14).

La traduzione automatica si inizia a sviluppare fin dagli anni '30, ma l'interesse cresce maggiormente a partire dalla fine degli anni '40 grazie al ricercatore statunitense Warren Weaver, che proprio nel 1949 pubblica il *Weaver Memorandum*, considerato come il documento che dà inizio alla ricerca nel campo della traduzione automatica. Nel suo Memorandum Weaver espone alcune problematiche legate allo sviluppo della medesima, ma presenta in maniera ottimista le capacità che i sistemi potrebbero

² Tutte le traduzioni di questo elaborato sono dell'autrice, salvo dove diversamente specificato.

raggiungere. Il suo Memorandum suscita un notevole interesse sull'argomento e proprio grazie al suo lavoro vengono fondati alcuni gruppi di ricerca in materia sia negli Stati Uniti che in Europa.

Il 1952 è l'anno in cui per la prima volta viene organizzata una conferenza sulla TA e due anni dopo viene presentato per la prima volta un sistema di TA tra il russo e l'inglese, sviluppato dall'azienda IBM in collaborazione con la Georgetown University. Nello specifico, il sistema era in grado di tradurre un insieme ristretto di frasi servendosi di un vocabolario di 250 parole e solo sei regole grammaticali. La dimostrazione portò allo stanziamento di ingenti finanziamenti in questo campo, principalmente dagli Stati Uniti e dall'ex Unione Sovietica. Negli anni '50 gli investimenti proseguirono e nel 1954 la TA fu oggetto per la prima volta di una conferenza internazionale e nello stesso anno venne creata la rivista *Mechanical Translation*. L'approccio alla base dei sistemi utilizzati era diretto, ovvero parola per parola, e l'obiettivo principale era quello di raggiungere un paradigma chiamato *Fully Automated High Quality Machine Translation* (FAHQMT). Lo scopo era quindi quello di creare un sistema in grado di produrre traduzioni in modo totalmente automatico e di alta qualità. Questo intento era frutto dell'ottimismo che aveva caratterizzato la ricerca sin dall'inizio; tuttavia nonostante i numerosi sforzi nessun sistema era ancora in grado di fornire traduzioni di qualità pari a quella umana, come riportato anche da Hutchins e Somers per i quali "optimism had been high, there were many predictions of imminent breakthroughs, but disillusionment grew as the complexity of the linguistic problems became more and more apparent" (Hutchins e Somers 1992:6). Dopo l'iniziale ondata di ottimismo i ricercatori si resero dunque ben presto conto della complessità dei problemi linguistici insiti nello sviluppo di sistemi di traduzione automatica.

Nel 1960 il primo a criticare il concetto di FAHQMT è stato lo scienziato statunitense Yeoshua Bar-Hillel, che lo definisce come un obiettivo irrealistico per la TA, a causa di numerose variabili, tra le quali la mancanza di conoscenza enciclopedica da parte dei sistemi, che li rendevano totalmente incapaci di risolvere i problemi di ambiguità semantica del linguaggio. Il periodo di grande entusiasmo nei confronti della TA si concluse nel 1964 con la creazione dell'*Automatic Language Processing Advisory Committee* (ALPAC), un ente incaricato di stilare un report sugli avanzamenti della ricerca e sulle prospettive della TA.

L'ente stilò l'ALPAC Report, un documento che espone tre limiti principali della traduzione automatica: la mancanza di una domanda di traduzione così alta da giustificare l'impiego della TA, l'incapacità di produrre una riduzione effettiva dei costi e dei tempi di traduzione e l'impossibilità per i sistemi di produrre in maniera totalmente automatica un output utilizzabile di testi generici. Con queste osservazioni viene quindi sancita la fine degli ingenti finanziamenti pubblici verso la TA negli Stati Uniti, ma nel decennio successivo la ricerca, con finalità e obiettivi più realistici, continua a svilupparsi in Canada e in Europa, dove la domanda di traduzione era in crescente aumento.

Negli anni '70 la ricerca si concentra sull'uso della traduzione automatica solo per alcune tipologie testuali e con dei domini specifici. Seguendo questi principi viene sviluppato in Canada il primo sistema che utilizzava il concetto di *sublanguage*, che possiamo definire come la lingua di uno specifico ambito tecnico, con un dizionario limitato e strutture sintattiche fisse. Il sistema, chiamato Météo, permetteva di tradurre i bollettini metereologici dall'inglese al francese con risultati eccellenti.

L'interesse per la TA non è però confinato solo al Canada, infatti anche nella Comunità Economica Europea (CEE) la domanda sempre crescente di traduzioni nelle varie lingue della comunità porterà all'acquisto da parte della CEE del sistema di TA sviluppato dall'azienda Systran. Nel 1978 viene inoltre finanziato un progetto per la realizzazione di un sistema in grado di tradurre tra tutte le lingue della CEE, chiamato Eurotra.

Negli anni '80 si assiste a un continuo progresso nella ricerca e allo stesso tempo compaiono per la prima volta anche i sistemi commerciali di traduzione automatica, diretti non più solo a istituzioni o enti di ricerca ma anche ad aziende, come per esempio i sistemi sviluppati da Systran.

Durante i primi anni '90 i sistemi commerciali continuano a diffondersi e in particolare si nota l'impegno in questo campo di aziende come IBM, Siemens e Microsoft. La prima delle tre sviluppa ben cinque sistemi di traduzione basati su dati statistici, un approccio innovativo che migliorerà l'output delle traduzioni e che sarà impiegato come metodo principale per alla TA per molti anni. Vengono inoltre impiegati sistemi basati su corpora di traduzioni, con un metodo che verrà in seguito definito *example-based*.

Gli anni 2000 sono caratterizzati dalla crescita dell'area della localizzazione software e dei CAT Tool, diretta conseguenza della diffusione dei PC. Negli anni 2000 si assiste

all'avvento di Internet, che aumenta in maniera significativa il numero potenziale di utenti della TA, sempre più spesso disponibile gratuitamente. Proprio in questo periodo vengono lanciati alcuni sistemi di traduzione automatica gratuita, tra cui Babel Fish nel 2003 (diventato poi Bing Translator nel 2009), Google Translate nel 2006 o Yandex nel 2011.

A partire dalla seconda decade del 2000 sono state implementate nuove tecnologie di apprendimento automatico, come per esempio l'apprendimento profondo o le reti neurali, elementi che ben presto si sono dimostrati adatti a diverse attività complesse, tra cui la traduzione automatica.

Nel 2014 viene pubblicato il primo articolo scientifico riguardante l'uso di reti neurali per la traduzione automatica (Bahdanau et al. 2014) e nel 2015 viene per la prima volta utilizzato un sistema di traduzione automatica neurale in una competizione sulla TA (OpenMT15).

Nel giro di pochi anni l'attenzione della ricerca si è rivolta principalmente all'architettura neurale. Basti pensare che alla *shared task* per la traduzione automatica organizzata dalla *Conference on Empirical Methods on Natural Language Processing* nel 2015 era stato presentato un solo sistema di traduzione neurale: i risultati erano stati buoni, ma non avevano eguagliato i ben più rodati sistemi statistici. Un anno dopo, allo stesso evento, un sistema di traduzione automatica neurale ha vinto in quasi tutte le coppie linguistiche. Nel 2017, quasi tutti gli iscritti hanno partecipato presentando un sistema di traduzione automatica neurale (Koehn, 2017:6).

Le architetture neurali hanno ormai surclassato in termini di qualità e velocità i sistemi statistici: per esempio nel 2016 Google ha annunciato il passaggio alle reti neurali per il suo sistema di traduzione automatica Google Translate, disponibile in più di 100 lingue ed utilizzato da più di 200 milioni di utenti ogni giorno.

1.3 Architetture dei sistemi di TA

1.3.1 Premessa

Nella sezione precedente sono stati esposti brevemente le origini e i principali sviluppi della traduzione automatica, mentre in queste sezioni verranno illustrate le principali architetture utilizzate in passato e attualmente.

1.3.2 Architetture *rule-based*

I primi approcci alla traduzione automatica hanno utilizzato architetture definite *rule-based*.

Questo tipo di architettura si compone principalmente di due elementi, ovvero dizionari monolingue e bilingue e un *parser*. Questo approccio, il primo e più rudimentale fra quelli che verranno analizzati, presenta numerosi limiti. Primo fra tutti è il tempo impiegato nella fase di *processing*, la mancanza di analisi del testo di partenza e l'incapacità da parte del sistema di risolvere aspetti ambigui della lingua, come per esempio di tradurre correttamente metafore. I due approcci principali di questa architettura sono definiti *interlingua* e *transfer*.

Nell'approccio *interlingua* il testo di partenza viene convertito in una rappresentazione astratta che esprime le informazioni sintattiche e semantiche del testo di partenza per poi convertirlo nel testo di arrivo.

Il secondo metodo è invece quello denominato *transfer*, nel quale il sistema viene implementato con un dizionario per la lingua di partenza, un dizionario bilingue e un dizionario per la lingua di arrivo. In questo approccio il testo di partenza viene convertito in una rappresentazione astratta, successivamente trasferita nella rappresentazione astratta del testo di arrivo per poi generare il testo nella lingua di arrivo.

1.3.3 Architetture statistiche

Le architetture statistiche e basate sui corpora si diffondono maggiormente dalla fine degli anni '90 e hanno rappresentato un grande passo avanti nella storia della traduzione automatica. Le architetture statistiche si basano sulle tecniche di calcolo probabilistico e sull'integrazione fra la conoscenza di eventi precedenti (in questo caso di traduzioni precedenti) con nuovi dati (in questo caso nuovi testi da tradurre). Nelle architetture statistiche il testo di partenza viene per prima cosa suddiviso in stringhe di parole e frasi, denominati segmenti. I segmenti vengono poi confrontati con un corpus bilingue allineato e a questo punto un metodo di calcolo statistico viene impiegato per ottenere nuovi segmenti nella lingua di arrivo, generando infine un testo nella lingua di arrivo.

L'architettura statistica è composta da un algoritmo che, tramite un modello di traduzione e un modello di lingua, individua quale traduzione ha la maggiore probabilità di essere “corretta”.

Il modello di traduzione si può basare su parole, frasi o insiemi di parole, tuttavia solitamente in questi tipi di sistemi di TA il riferimento è dato dalle singole parole. Il modello di lingua invece è basato su n-grammi, cioè su una sequenza di elementi (parole o lettere in questo caso) e sul calcolo della probabilità di una determinata sequenza di parole nel testo di arrivo.

Un altro sviluppo di questa architettura è l'approccio basato su esempi o *example-based*. Questo sistema solitamente opera in tre fasi. Per prima cosa il sistema cerca di identificare segmenti del nuovo testo nella lingua di partenza con esempi (ovvero sequenze esatte di parole) estratte dal corpus bilingue allineato e in questa fase tutti i frammenti pertinenti vengono raccolti e conservati. Successivamente il sistema allinea i segmenti corrispondenti alle traduzioni utilizzando gli esempi del corpus e infine combina i segmenti per generare il testo nella lingua di arrivo.

1.3.4 Architetture neurali

A partire dal 2014 è emersa una nuova architettura, la traduzione automatica neurale, che attualmente è impiegata non solo nei sistemi più all'avanguardia, ma da tutti i principali sistemi di commerciali, sia a pagamento che gratuitamente disponibili sul web.

La TA neurale è basata sul *deep learning* (apprendimento profondo) una tecnica che permette, tramite la creazione di reti neurali artificiali, di apprendere e svolgere una determinata attività in autonomia da parte del sistema.

Il primo neurone artificiale è stato sviluppato nel 1943 da McCulloch e Pitts e questa tipologia di approccio è tutt'ora utilizzata nelle reti neurali. Le reti neurali artificiali sono state sviluppate cercando di imitare il funzionamento dei neuroni del cervello umano e operano, in maniera molto semplificata, proprio come i neuroni umani.

Un neurone riceve un segnale di input che viene processato tramite degli algoritmi e viene successivamente “inviato” ad un altro neurone, per poi produrre un output finale. Hecht-Nielsen definisce infatti le reti neurali come “a computing system made up of a number of simple, highly interconnected processing elements, which process

information by their dynamic state response to external inputs” (Hecht-Nielson in Caudill, 1989).

Il *deep learning* opera solitamente partendo da un numero molto ampio di esempi da cui estrapola automaticamente le caratteristiche principali, apprendendole in maniera gerarchica, partendo quindi da elementi di base, fino a identificare strutture più complesse e ottenendo infine un’analisi generale del fenomeno di studio. Possiamo dire che l’apprendimento profondo imita il tipo di apprendimento di un bambino: partendo dall’osservazione e dagli stimoli esterni riesce a ricavare regole e apprendere delle informazioni. Allo stesso modo le reti neurali hanno come input quantità molto ampie di dati che permettono al sistema di trovare dei *pattern* e di calcolare delle probabilità in modo di riuscire ad “apprendere” in tempi rapidi.

Gli elementi di elaborazione della rete vengono definiti nodi o neuroni e vengono solitamente dislocati in vari livelli. Il primo livello, chiamato livello di ingresso o *input layer* riceve le informazioni ottenute dall’esterno e le processa, mentre il livello di uscita o *output layer* mostra il risultato di quanto il sistema è riuscito a imparare. Fra questi due livelli si trovano uno o più livelli nascosti o *hidden layers*, che non hanno contatto diretto con l’esterno e dove vengono effettuati i calcoli necessari all’apprendimento.

I sistemi di traduzione automatica neurale (NMT) sono solitamente formati da un *encoder* che trasforma l’input di partenza (per esempio una frase o un testo) in una lista di vettori, ovvero una sequenza di numeri che rappresenta il significato del testo di partenza. I vettori vengono a loro volta decodificati tramite un *decoder* per ottenere l’output nella lingua di arrivo. I sistemi di traduzione automatica neurale attribuiscono a ogni parola un vettore, “posizionandolo” in una precisa dimensione, permettendo quindi di rappresentare numerose informazioni riguardanti la parola e il contesto in cui appare, infatti “the feature vector represents different aspects of the word: each word is associated with a point in a vector space” (Bengio et al., 2003: 1139).

Riprendendo le parole di Firth (“*you shall know a word by the company it keeps*”, Firth, 1957:1) possiamo comprendere il concetto di *word embedding*. Questo processo, esposto precedentemente, permette di collocare in uno spazio multidimensionale le parole, mostrando spazialmente la vicinanza che hanno le parole con caratteristiche in comune. Per esempio le parole che appartengono alla stessa categoria grammaticale verranno raggruppate in una dimensione, le parole che condividono lo stesso genere grammaticale in un’altra, etc.

Meaning and semantics are quite difficult concepts with largely unresolved definition. The idea of distributional lexical semantics is to define word meaning by their distributional properties, i.e., in which contexts they occur. Words that occur in similar contexts (dog and cat) should have similar representations. In vector space models, such as the word embeddings that we use here, similarity can be measured by a distance function, e.g., the cosine distance—the angle between the vectors. (Koehn 2017:36)

Le tipologie di reti neurali principalmente utilizzate nella traduzione automatica neurale sono le reti neurali ricorrenti (RNN) e le reti neurali convoluzionali (CNN). Le prime sono utilizzate per l'elaborazione del linguaggio naturale, mentre le seconde sono state solitamente utilizzate per il riconoscimento di immagini, ma da alcuni anni vengono impiegate anche nella creazione di sistemi di TA.

1.3.4.1 Reti Neurali Ricorrenti (RNN)

Le RNN sono modelli specializzati per i dati sequenziali. Si tratta di componenti di rete che prendono come input una sequenza di elementi e producono un vettore di dimensione fissa che riassume tale sequenza (Goldberg 2017:3).

Le RNN processano il linguaggio in maniera sequenziale seguendo la direzione sinistra verso destra o viceversa, analizzando una parola per volta, e questo porta il sistema ad eseguire più passaggi per prendere decisioni che dipendono da parole lontane l'una dall'altra (Uszkoreit, 2017: online). Come precedentemente esposto, le reti neurali di solito processano la lingua generando rappresentazioni vettoriali spaziali e successivamente aggregano le informazioni dalle parole circostanti per determinare il significato di una parola nel suo contesto.

Utilizzando l'esempio fornito dal ricercatore di Google Jakob Uszkoreit nella frase “I arrived at the bank after crossing the...” (dove bank può significare sia banca che riva) la decisione sul significato della parola bank e sulla sua conseguente rappresentazione vettoriale può arrivare solo sapendo se la frase termina con “...road” ossia strada, o con “...river” ossia fiume. Rispetto all'esempio riportato, una RNN può determinare il significato del termine bank solo dopo aver analizzato ogni parola tra “bank” e “river” (*ibid.*) e, come dimostrato anche dalla ricerca di Bengio et al. (2003), più

passaggi sono richiesti per effettuare questa scelta, più è difficile per il sistema apprendere come effettuarla correttamente.

1.3.4.2 Reti Neurali Convoluzionali (CNN)

Un altro sistema attualmente utilizzato nella traduzione automatica è la rete neurale convoluzionale (CNN). Le CNN sono architetture che solo recentemente sono state utilizzate nella traduzione, in quanto sono particolarmente efficaci nell'estrazione di *pattern* locali a partire da un numero molto alto di dati e la loro prima applicazione è stata quella del riconoscimento di immagini.

Queste reti sono in grado di estrarre modelli locali sensibili all'ordine delle parole, indipendentemente da dove appaiano nell'input, permettendo al sistema di identificare espressioni idiomatiche fino a una lunghezza fissa in frasi più lunghe o in interi documenti (Goldberg 2017:3).

Una rete neurale convoluzionale permette di applicare le previsioni elaborate dal sistema su n-grammi che hanno caratteristiche comuni anche a n-grammi mai incontrati dal sistema stesso. Nel 2017 i ricercatori di Facebook Artificial Intelligence Research (FAIR) hanno introdotto l'utilizzo delle reti neurali convoluzionali nei sistemi di traduzioni automatica dell'omonimo social network.

Nonostante le RNN abbiano storicamente superato le CNN nelle attività di traduzione, occorre comprendere il modo in cui elaborano le informazioni per capire quali limiti questa struttura possa avere. Le RNN operano leggendo l'input da sinistra a destra o viceversa, una parola alla volta e per questo motivo il sistema non può effettuare dei calcoli in maniera parallela, poiché prima di analizzare una nuova parola deve terminare l'analisi di quella precedente. Viceversa le CNN riescono a calcolare tutti gli elementi di una frase simultaneamente, e sono quindi più efficienti dal punto di vista computazionale (Gehring, 2017: online). Un altro vantaggio delle CNN è che le informazioni sono elaborate gerarchicamente, rendendo più facile catturare relazioni complesse nei dati.

1.3.5 Intelligenza artificiale e società

Lo sviluppo dell'intelligenza artificiale (*artificial intelligence* o *AI*) e delle reti neurali ha portato ad un utilizzo sempre più elevato degli algoritmi nella vita quotidiana di milioni di persone. Sono infatti sempre più numerosi gli impieghi dell'intelligenza

artificiale, dal riconoscimento facciale per sbloccare lo schermo dello smartphone all'analisi automatica dei CV per un impiego, fino ai sistemi utilizzati in campo medico per fornire diagnosi ai pazienti. L'AI è quindi diventata uno strumento utilissimo e rilevante nel lavoro di molte persone, e la ricerca produce nuove applicazioni sempre più precise a ritmi elevatissimi. Negli ultimi anni, tuttavia, la comunità scientifica ha iniziato una riflessione riguardo l'equità e l'imparzialità dell'intelligenza artificiale, in particolare nel campo dell'elaborazione del linguaggio naturale (NLP). Inoltre, l'elaborazione del linguaggio naturale (e di conseguenza anche uno dei suoi campi di applicazione, la traduzione automatica) viene impiegata in un numero sempre maggiore di processi, basti pensare alla traduzione automatica presente ormai in tutti i maggiori social network o alla generazione automatica di sottotitoli nei video offerta dalla piattaforma Youtube. Per questo motivo di seguito verranno esposte alcune nuove applicazioni della traduzione automatica che hanno un impatto diretto sulla società e alcuni studi che prendono in considerazione le criticità legate all'uso di sistemi di NLP, in particolare la presenza (e il perpetuarsi) di stereotipi di genere nell'elaborazione del linguaggio naturale, con una particolare attenzione alla TA.

1.3.5.1 Nuove applicazioni per la traduzione automatica

La traduzione automatica e l'elaborazione del linguaggio naturale (NLP) sono campi in cui la ricerca sta progredendo ad una velocità elevatissima, spaziando dal possibile utilizzo della traduzione automatica per testi letterari (si veda per esempio Toral et al., 2018) agli studi sulla presenza di stereotipi di genere nell'elaborazione del linguaggio naturale e in particolare nella traduzione automatica (Costa-jussà, 2019), argomento a cui verrà dedicato una sezione specifica in questo elaborato. I campi di applicazione di questa nuova tecnologia sono numerosi e in questa sezione mi concentrerò solo su alcuni progetti che hanno un impatto sulla società.

Da quanto visto nelle sezioni precedenti, si può notare come la TA sia stata per lungo tempo oggetto di studio e di interesse principalmente da parte del mondo accademico e dalle aziende, tralasciando il ruolo sociale che questa tecnologia può svolgere. Uno degli obiettivi di questo elaborato è invece quello di presentare lo studio della TA non solo in un'ottica di ricerca, ma di comprendere i risvolti che l'uso di questa tecnologia ha sulla società e in particolare sugli utenti che usufruiscono dei servizi di traduzione

offerti dai numerosi sistemi automatici online. Ora che la qualità dell'output prodotto dai sistemi di traduzione automatica è decisamente migliorato (nonostante sia ancora in molti casi lontano dai livelli della traduzione umana) la TA può infatti essere vista come uno strumento che può essere messo al servizio della società.

L'interesse verso la ricerca in questo campo è dimostrata per esempio dalla presenza di progetti riguardanti TA e NLP tra i progetti finanziati dall'Unione Europea nel programma di ricerca e innovazione denominato Horizon 2020, per il quale sono stati stanziati complessivamente quasi 80 miliardi di euro nel periodo 2014-2020. Tra i progetti finanziati di particolare rilevanza sono il progetto *Health in my language* (HimL) e *International Network on Crisis Translation* (Interact).

Il progetto "*Health in my Language*", coordinato da Barry Haddow dell'University of Edinburgh, è stato attivo dal 2015 al 2018. HimL mirava a rendere le informazioni sanitarie contenute in specifiche pagine web disponibili in una più ampia varietà di lingue (ceco, tedesco, polacco e rumeno) utilizzando la traduzione automatica. L'obiettivo di questo progetto è stato quindi quello di superare le barriere linguistiche permettendo a un numero più alto di utenti di poter accedere alle informazioni sanitarie presenti nei siti tradotti. Il progetto assume rilevanza poiché permette da un lato di rendere fruibili più informazioni per i cittadini e dall'altro di creare risorse che potranno essere utilizzate in futuro per la creazione di altri sistemi di traduzione automatica per lingue che hanno scarse risorse.

Un altro progetto finanziato nell'ambito di Horizon 2020 è il progetto *Interact*, coordinato da Sharon O'Brien, professoressa dell'Università di Dublin City. Tra i vari obiettivi e campi di applicazione della ricerca, vi è lo sviluppo e l'uso di sistemi di traduzione automatica in contesti di emergenza e crisi. Il progetto, ancora in corso, definisce i contesti di azione e si concentra sullo sviluppo di sistemi di TA per lingue con poche risorse, sulla sensibilizzazione di traduttori professionisti riguardo all'uso di questi strumenti e sull'informazione del cittadino riguardo i sistemi e come poterli utilizzare nella maniera migliore in un contesto di crisi. Questa ricerca è particolarmente rilevante per l'impatto sociale che propone, dato che si basa sulla volontà di considerare l'accesso alle informazioni come un diritto di tutti i cittadini, in maniera particolare durante le emergenze. Come sottolineato dai membri del progetto, la comunicazione durante una crisi deve essere multilingue e multiculturale, in modo da potersi adeguare alle esigenze di una società sempre più globalizzata, con una

crescente urbanizzazione e migrazioni. Per questi motivi, la comunicazione multilingue e multiculturale può essere raggiunta tramite la traduzione³.

Infine, un altro aspetto che ha un impatto sulla società è quello che interessa l'ecosostenibilità legata allo sviluppo e al costante miglioramento di sistemi di traduzione automatica neurale sempre più sofisticati. Alla luce dello studio condotto da Strubell et al. (2019), che dimostra gli immensi consumi energetici legati allo sviluppo delle reti neurali, occorre porre l'attenzione non solo sulle ripercussioni che l'uso dell'intelligenza artificiale (e di conseguenza anche della TA) può avere su aspetti rilevanti della società come il perpetuare o ingigantire stereotipi di genere o razziali, ma anche sugli effetti che queste tecnologie hanno a livello ambientale. Come auspicato da Strubell et al., “we recommend a concerted effort by industry and academia to promote research of more computationally efficient algorithms, as well as hardware that requires less energy” (2019: 3649). Emerge quindi il bisogno di iniziare a creare sistemi di TA efficienti non solo dal punto di vista linguistico, ma anche da quello della sostenibilità.

1.3.6 Mind the gap: il gap di genere nell'elaborazione del linguaggio naturale

I recenti sviluppi delle applicazioni di *machine learning*, come per esempio le reti neurali profonde, hanno rivoluzionato molti processi nell'elaborazione del linguaggio naturale, anche grazie alla disponibilità sempre maggiore di dati utilizzabili per “addestrare” i vari sistemi. Tuttavia sempre più studiosi si stanno accorgendo che i sistemi non solo riproducono gli stereotipi presenti nei set utilizzati per l'addestramento, ma rischiano addirittura di amplificare questi stereotipi, infatti “learning algorithms run the risk of relying on societal biases encoded in the training data to make predictions” (Chang et al., 2019). Questo fenomeno è definito anche come *algorithmic bias*, ovvero la presenza nei data set o nei software stessi di stereotipi, infatti “if the input data is indeed biased, the output of the algorithm might also reflect the same bias” (Baeza-Yates, 2018: 59).

In particolare un nuovo filone di ricerca si sta occupando dello studio degli stereotipi di genere nell'elaborazione del linguaggio naturale e dei possibili metodi per mitigare o eliminare gli stereotipi presenti. L'interesse verso questo ambito viene dimostrato anche dal crescente numero di seminari e workshop sul tema, come per esempio il

³ <https://cordis.europa.eu/project/id/734211>

First International Workshop on Gender equality in Software Engineering (ACM/IEEE maggio 2018) o il *First Workshop on Gender Bias in Natural Language Processing* (ACL agosto 2019). Il tema è di particolare rilevanza anche perché “la sovra rappresentazione degli uomini nella creazione di queste tecnologie potrebbe tranquillamente cancellare decenni di conquiste per l’uguaglianza di genere” (Leavy, 2019:14) nonché esporre quello che la scrittrice Caroline Criado Perez definisce come “*gender data gap*” (2019: XII) ovvero l’assenza di un’adeguata rappresentazione delle donne nei set statistici. Le donne infatti non sono solo meno rappresentate, ma gli studi che verranno presentati nelle sezioni successive dimostrano una forte tendenza all’amplificazione di stereotipi di genere acquisiti tramite i dati di *training*.

1.3.6.1 Alcuni esempi di studi sugli stereotipi di genere nel NLP

Come precedentemente esposto, l’interesse verso il riconoscimento e l’eliminazione degli stereotipi di genere nell’elaborazione del linguaggio naturale è un argomento che viene affrontato sempre più frequentemente dagli studiosi di vari campi. In questa sezione verranno presentati alcuni studi sulla presenza e le varie strategie per mitigare o eliminare gli stereotipi in diversi campi di applicazione del NLP, come per esempio la *sentiment analysis* o la risoluzione della coreferenza (*coreference o anaphora resolution*), mentre si rimanda alla sezione 1.3.6.2 di questo elaborato per la presentazione degli studi che riguardano la traduzione automatica.

Bhaskaran e Bhallamudi (2019) affrontano la questione degli stereotipi legati alle occupazioni nella *sentiment analysis*, con l’obiettivo di misurare la differenza statistica tra coppie di frasi contenenti un referente maschile o femminile e una professione. Per l’analisi è stato creato un set composto da 800 frasi con la struttura “<noun> is a/an <profession>” (Bhaskaran e Bhallamudi, 2019:63) dove *noun* corrisponde a un sintagma nominale maschile o femminile e *profession* è una professione tra le 20 selezionate per lo studio. Il set è bilanciato tra il genere maschile e femminile, poiché per ogni professione sono presenti 40 coppie di sintagmi nominali, 20 con referente femminile e 20 con referente maschile, ottenendo quindi coppie di frasi come “He is a doctor/she is a doctor” (*ibid.*). I risultati dello studio evidenziano come nei modelli utilizzati siano presenti stereotipi di genere, mostrando “a statistically significant difference between the two genders” (Bhaskaran e Bhallamudi, 2019: 65) nell’attribuzione di sentiment positivi o negative a frasi contenenti riferimenti maschili

o femminili. Come emerge dallo studio, “biases present in training data can get propagated through machine learning models” (*ibid.*) e di conseguenza l’aumento degli stereotipi potrebbe causare discriminazioni dato l’uso sempre più diffuso di modelli di *sentiment analysis* al fine di valutare automaticamente i curricula per una posizione di lavoro (Bhaskaran e Bhallamudi, 2019:66).

Un altro filone di studi particolarmente rilevanti è rappresentato dagli studi riguardanti la risoluzione della coreferenza, ovvero l’utilizzo di algoritmi per decifrare “espressioni che si riferiscono a una stessa entità nel testo.”⁴ Lo studio di Rudinger et al. (2018) si propone di analizzare la presenza di stereotipi di genere nei sistemi utilizzati per “risolvere” la coreferenza attraverso la creazione di un set *ad hoc* basato sullo schema di Winograd (Levesque et al., 2012). Il suddetto schema, elaborato originariamente in lingua inglese è composto da una coppia di frasi che si differenziano tra di loro solo per una parola e contengono un pronome che, nelle due frasi, si riferisce a due diversi antecedenti. Per poter disambiguare il referente corretto della frase è pertanto richiesta una conoscenza del mondo reale, rendendo il compito facile per gli esseri umani ma complicato per i sistemi *data-driven* (Levesque et al., 2012). Il set elaborato da Rudinger et al., chiamato “Winogender” ha una struttura frasale che comprende un’occupazione, “a person referred to by their occupation and a definite article, e.g. the paramedic” (Rudinger et al. 2018:9), un partecipante, “a secondary (human) participant, e.g. the passenger (*ibid.*) e un pronome, coreferente o dell’occupazione o del partecipante della frase. Per lo studio sono state selezionate 60 professioni con le relative percentuali di occupazione divise per genere selezionate dal *Bureau of Labor Statistics* degli Stati Uniti. Il set comprende un totale di 720 frasi che hanno ricevuto sia una valutazione umana che automatica, con l’obiettivo di individuare i casi in cui il sistema è più o meno incline a riconoscere un pronome come coreferente a seconda delle diverse professioni. I risultati dello studio mettono in luce come siano presenti degli stereotipi relativi alle professioni nella risoluzione della coreferenza, in quanto gli algoritmi tendono a “un’eccessiva generalizzazione degli attributi del genere” (Rudinger et al., 2018:12), mostrando per esempio come i sistemi peggiorino notevolmente in “sentences in which the pronoun gender does not match the occupation’s majority gender” (Rudinger et al., 2018: 11).

⁴ Definizione elaborata dallo *Stanford Natural Language Processing Group* (traduzione mia): <https://nlp.stanford.edu/projects/coref.shtml>

Nello stesso periodo è stato condotto anche un altro studio sulla coreferenza che si concentra sugli stereotipi di genere riguardanti le professioni (Zhao et al. 2018), che dimostra come vari sistemi elaborati per risolvere l'ambiguità della coreferenza (basati su regole, statistici e neurali) mostrano una preferenza nell'attribuzione del pronome riferito a occupazioni stereotipicamente maschili o femminili. Il set realizzato per lo studio del fenomeno è denominato "Winobias" (Zhao et al., 2018: 15) ed è composto da frasi basate sullo schema di Winograd contenenti 40 professioni selezionate dal *Bureau of Labor Statistics* degli Stati Uniti, ottenendo un totale di 3160 frasi. La struttura frasale dei set è così descritta:

It contains two types of challenge sentences that require linking gendered pronouns to either male or female stereotypical occupations. None of the examples can be disambiguated by the gender of the pronoun but this cue can potentially distract the model. We consider a system to be gender biased if it links pronouns to occupations dominated by the gender of the pronoun (pro-stereotyped condition) more accurately than occupations not dominated by the gender of the pronoun (anti-stereotyped condition). (Zhao et al., 2018:16)

Attraverso l'analisi del corpus di addestramento del sistema, i ricercatori hanno evidenziato come "le entità femminili sono significativamente sotto-rappresentate" (Zhao et al., 2018:16), quindi per cercare di arginare l'impatto di tali distorsioni "viene proposto di generare un dataset ausiliario in cui tutti i referenti maschili siano sostituiti da referenti femminili e viceversa" (*ibid*). Lo studio infine dimostra la presenza di stereotipi di genere nella risoluzione della coreferenza, mostrando inoltre che "gli stereotipi nei sistemi di NLP hanno il potenziale non solo di mostrare, ma anche di amplificare gli stereotipi presenti nella società" (Zhao et. al., 2018:19).

1.3.6.2 Stereotipi di genere e traduzione automatica

L'ambito della traduzione automatica, come alcuni altri campi di applicazione dell'elaborazione del linguaggio naturale, negli ultimi anni ha iniziato interrogarsi su molte criticità legate all'uso di tali sistemi, individuando negli stereotipi di genere uno degli argomenti da analizzare e affrontare. Nel 2018 Google ha annunciato l'introduzione della traduzione maschile e femminile per i singoli termini in alcune lingue, tra cui l'italiano, quando si parte da una fonte inglese; con l'obiettivo di ridurre gli stereotipi di genere nel suo sistema di traduzione automatica Google Translate (Kuczmarski, 2018). Questo primo passo dimostra l'interesse crescente verso

l'argomento, e in questa sezione verranno esposti alcuni studi sull'argomento in varie combinazioni linguistiche che propongono diversi metodi per mitigare gli stereotipi. Uno dei primi studi è quello di Prates et al. (2018) in cui la traduzione automatica viene utilizzata come strumento attraverso il quale studiare il fenomeno degli stereotipi di genere nell'intelligenza artificiale (Prates et al., 2018:1). Per lo studio è stato elaborato un set contenente più di mille professioni selezionate dal *Bureau of Labor Statistics* degli Stati Uniti, tradotto da 12 lingue che non hanno marcatura di genere verso la lingua inglese, con una struttura frasale che comprende un pronome neutro e una professione. È stato inoltre sviluppato un secondo set di frasi contenenti 21 aggettivi che si trovano tra le prime mille parole più frequenti del *Corpus of Contemporary American English* (COCA). Lo studio, condotto ad aprile 2018, dimostra come il sistema tenda a utilizzare più frequentemente pronomi maschili rispetto a quelli femminili o neutri, inoltre “i risultati sembrano suggerire che questo fenomeno si estenda oltre l'ambito lavorativo, dato che la proporzione di pronomi femminili varia in modo significativo a seconda degli aggettivi usati per descrivere una persona” (Prates et al., 2018:28). Come in altri studi già presentati in questo elaborato, il sistema tende a sottostimare la frequenza dei pronomi femminili anche in professioni a maggioranza femminile, creando quindi un'asimmetria tra i pronomi femminili e maschili nell'output (*ibid.*)

Stanovsky et al. (2019) affronta la questione degli stereotipi di genere utilizzando la TA, ma partendo dai dataset elaborati da Rudinger et al (2018) e Zhao et al (2018) per gli studi sulla coreferenza, ossia “Winogender” e “Winobias”. Il set, creato dall'unione dei due set precedenti, è stato chiamato “WinoMT” e contiene 3888 frasi bilanciate sia fra genere femminile e maschile che fra ruoli stereotipici e non. Il set è stato successivamente tradotto automaticamente in 8 lingue che presentano genere grammaticale (tra le quali anche l'italiano) utilizzando 6 sistemi di TA differenti, 4 commerciali (tra cui Google Translate) e 2 usati nella ricerca accademica. I risultati dello studio hanno dimostrato che tutti i sistemi presentano stereotipi di genere e inoltre

the systems have a significant and consistently better performance when presented with pro-stereotypical assignments (eg. female nurse) while their performance deteriorates when translating anti-stereotypical roles (e.g. male receptionist) (Stanovsky et al., 2019:1682).

È stato inoltre creato un secondo set per verificare l'ipotesi che alcuni aggettivi solitamente riferiti a uomini o donne possano modificare l'output della traduzione. A questo scopo al set originale sono stati aggiunti gli aggettivi *handsome* e *pretty*, che sono solitamente utilizzati per riferirsi rispettivamente a uomini o donne, con l'obiettivo di valutare se questa aggiunta potesse modificare lo "stereotipo" contenuto nella professione "mischiando" i segnali. Infatti, mentre il sostantivo *doctor* farebbe propendere verso una traduzione maschile, l'aggettivo *pretty* farebbe propendere la traduzione verso la forma femminile (*ibid.*). I risultati del test hanno dimostrato che questo metodo ha migliorato l'output in alcune combinazioni linguistiche, tuttavia viene sottolineato come non possa essere utilizzato come metodo per mitigare la presenza degli stereotipi di genere, in primo luogo perché si basa su aggettivi stereotipici di per sé e in secondo luogo perché sarebbe impossibile da implementare in data set molto ampi.

L'ultimo contributo che verrà presentato in questa sezione è quello elaborato da Escudé Font e Costa-jussà (2019) il cui obiettivo è l'individuazione e lo studio di stereotipi di genere nella TA e la proposta di utilizzare i *word embeddings* per mitigarne l'impatto. Lo studio è stato condotto utilizzando un sistema di TA creato *ad hoc* e, per testare la presenza di stereotipi di genere, è stato creato un set apposito contenente più mille professioni selezionate dal *Bureau of Labor Statistics* degli Stati Uniti. La struttura frasale scelta per il set è "*I've known her/him/proper noun for a long time, my friend works as a/an <occupation>*" (Escudé Font e Costa-jussà, 2019: 149) nella combinazione linguistica inglese-spagnolo. Nello studio i due autori hanno utilizzato nomi propri spagnoli per ridurre l'ambiguità, inoltre questa struttura frasale permette di valutare se la parola *friend* e le professioni vengono tradotte correttamente, ovvero al maschile o al femminile, a seconda del coreferente della frase. I risultati dello studio mettono in luce stereotipi di genere nel sistema, infatti il pronome maschile viene sempre identificato dal sistema, nonostante non tutte le occupazioni vengano tradotte correttamente (Escudé Font e Costa-jussà, 2019: 152), mentre il pronome femminile ha diverse percentuali di identificazione nei sistemi. Viene inoltre dimostrato che "the accuracy when predicting the gender for this test set is improved for some settings, when using the debiased and gender-neutral word embeddings." (*ibid.*) e che l'utilizzo di un metodo per mitigare la presenza degli stereotipi di genere tramite *word embeddings* mantiene la qualità dell'output. Lo studio propone di continuare la ricerca riguardante l'equità nei sistemi di TA e più in generale nelle reti

neurali, ponendo l'attenzione sulla possibile presenza di altri stereotipi nei *training data*, come per esempio quelli riferiti a etnia, credo religioso o età.

1.3.6.3 Stereotipi di genere e traduzione automatica: il focus sull'italiano

Gli studi presentati in questo capitolo si concentrano sull'individuazione degli stereotipi di genere in alcune applicazioni dell'elaborazione del linguaggio naturale e in diverse combinazioni linguistiche, tuttavia ad oggi non risultano presenti studi nella sola combinazione inglese-italiano, eccezion fatta per *Questioni di genere e traduzione automatica* di Monti (2017). Per questo motivo lo studio verrà presentato nel dettaglio, in quanto unico contributo per la combinazione linguistica che sarà oggetto di questo elaborato, nonché perché riesce a fornire un punto di vista legato non solo ad ambiti strettamente computazionali, ma anche linguistici e traduttivi.

Lo studio di Monti sottolinea come la questione del genere nella traduzione “benché ampiamente trattata e discussa nell’ambito della teoria della traduzione, in particolar modo di stampo femminista” (Monti, 2017:412) sia invece esaminata da un numero molto limitato di contributi riguardanti la traduzione automatica. I problemi legati alla traduzione del genere nella traduzione automatica

sono molteplici e di diversa natura, come la riproduzione di stereotipi di genere principalmente legati ai ruoli maschili e femminili nella società, la scorretta attribuzione del genere a proforme (pronomi personali, relativi, etc.), aggettivi, participi passati e apposizioni, oltre a un uso più frequente del genere maschile per pronomi e nomi anche quando questi si riferiscono al genere femminile. (*ibid.*)

Viene inoltre messa in evidenza la centralità del traduttore nella “corretta individuazione del genere naturale della persona a cui si riferisce il sostantivo”, in quanto potrebbe non corrispondere all’idea stereotipica che si associa a una professione o a un ruolo (Monti, 2017: 415), elemento che non deve essere trascurato quando vengono utilizzati sistemi di traduzione automatica. Al fine di valutare e confrontare la questione di genere nella TA Monti individua sei fenomeni linguistici che possono presentare difficoltà legate agli stereotipi di genere, ovvero:

1. Accordo soggetto - predicato nominale
2. Accordo soggetto|oggetto - complemento predicativo del soggetto|complemento predicativo dell'oggetto
3. Accordo nome - apposizione
4. Accordo nome - riferimento anaforico/cataforico

5. Accordo nome - participio passato
 6. Genere sociale
- (Monti 2017: 424)

dove per ogni fenomeno analizzato viene fornita una frase d'esempio, tradotta utilizzando sei diversi sistemi di traduzione automatica commerciali di diverse tipologie. Tra questi, tre sistemi sono di tipo neurale (seppure alcuni ancora in fase sperimentale), uno di tipo statistico, uno basato su regole e infine uno ibrido, che "coniuga l'approccio linguistico con quello statistico" (Monti, 2017: 423). Il risultato dello studio di Monti evidenzia come tutti i sistemi presi in analisi presentino degli stereotipi di genere, dimostrando che la questione non è relativa semplicemente ai sistemi neurali, ma anche ai sistemi statistici e *rule based*. Si può quindi ipotizzare che le criticità alla base di questa tematica siano legate principalmente ai dati utilizzati per la creazione e l'addestramento di sistemi, tanto statistici quanto neurali. La stessa Monti nel suo articolo evidenzia come sia necessario uno studio più esteso ed approfondito dell'argomento. Inoltre il suo contributo pone le basi per uno studio qualitativo che si proponga di analizzare il tipo di errore legato alla presenza di stereotipi di genere nei sistemi di TA, al fine di arrivare a "una comprensione più approfondita del fenomeno e dunque anche all'individuazione di possibili soluzioni per migliorare i sistemi attuali" (Monti, 2017: 431).

Capitolo II: genere e lingua

2.1. Introduzione

In questo capitolo verrà brevemente spiegato il concetto di stereotipo, con una particolare attenzione agli stereotipi di genere, si illustrerà inoltre il rapporto tra questioni di genere e lingua, concentrandosi in particolare sulla questione delle forme femminili di ruoli e professioni. Verrà quindi ripercorsa la storia legata al linguaggio di genere in Italia, esponendo il dibattito e la riflessione che alcune scelte linguistiche, come per esempio quello dei femminili professionali, hanno causato nella società e in particolare nella stampa e nella politica. Si esporranno poi gli atteggiamenti riguardo al femminile delle professioni in Francia e in Spagna, dal punto di vista delle politiche linguistiche intraprese e le posizioni dei rispettivi istituti per la salvaguardia della lingua francese e spagnola. Questi due paesi, infatti, seppure non oggetto diretto di studio in questo elaborato, presentano molte affinità linguistiche e culturali con l'italiano, e per questi motivi è stato ritenuto interessante fornire le prospettive riguardanti le azioni intraprese in questi paesi. Infine esporrò il caso della lingua inglese che, pur non presentando marcatura di genere per le professioni, ha attuato alcune politiche linguistiche per favorire l'uso di un linguaggio non sessista. L'obiettivo di questo capitolo è quindi di fornire una panoramica sul concetto di stereotipo, argomento che è stato anticipato nel capitolo I e che sarà oggetto dello studio di caso presentato nei capitoli III e IV, nonché di analizzare il rapporto tra la lingua e il genere nell'ambito dei sostantivi utilizzati per denominare professioni e incarichi.

2.2 Stereotipi e pregiudizi

Gli stereotipi possono essere visti come una semplificazione della comprensione del mondo, basati sulla categorizzazione che anche inconsapevolmente effettuiamo ogni giorno. Lo psicologo Henri Tajfel definisce in questa maniera lo stereotipo cognitivo, ovvero come un processo la cui funzione principale “consiste nel semplificare e sistematizzare [...] l'abbondanza e la complessità delle informazioni che l'essere umano riceve dall'ambiente” (1981, trad. it. 1995: 238). Questo tipo di stereotipi, nati dall'esigenza umana di classificare le persone che incontriamo nella nostra vita, le raggruppa in base alle caratteristiche socialmente rilevanti che condividono o agli

attributi distintivi che le differenziano. Aniché persone, attraverso gli stereotipi, diventano quindi uomini, donne, stranieri, anziani, etc. e in questo modo percepiamo un altro individuo non come unico e particolare, ma come membro di una categoria sociale. Tajfel aggiunge inoltre che lo stereotipo cognitivo si trasforma in uno stereotipo sociale nel momento in cui vengono “condivisi da grande masse di persone all’interno dei gruppi e istituzioni sociali” (*ibid.*)

Una prima definizione degli stereotipi è stata data dal giornalista Walter Lippmann, che li ha definiti “stampi cognitivi che riproducono le immagini mentali delle persone o gruppi” (1922), inoltre nella psicologia sociale gli stereotipi sono definiti come “le credenze sugli attributi personali di una categoria sociale” (Villano, 2003: 9). La creazione di stereotipi nasce quindi dall’esigenza di categorizzare le informazioni che possiamo dedurre da una persona e generalizzarle, inoltre si può affermare che

non è possibile accostarci agli altri in maniera del tutto libera da ipotesi e aspettative, delle quali abbiamo bisogno per poterci prefigurare il possibile svolgimento degli eventi [...] queste aspettative si esprimono molto spesso attraverso il ricorso a categorie sociali con relativi stereotipi, che dovrebbero essere poi sottoposte a verifica nel corso dell’interazione. (Mazzara 1997:100-101).

Come esposto da Bruno Mazzara, nel corso dell’interazione le ipotesi formulate a priori dovrebbero essere verificate e modificate, tuttavia è molto complesso cambiare idea riguardo a uno stereotipo. La modifica o l’eliminazione di uno stereotipo sono azioni difficili da compiere, sia perché sono pensieri che vengono interiorizzati sin dall’infanzia, sia perché sono condivisi dal nostro gruppo sociale e quindi largamente diffusi. Da quanto esposto emerge che gli stereotipi sono parte integrante della cultura di una certa società, costituendo vere e proprie “rappresentazioni sociali che si apprendono, si trasmettono e si evolvono tramite il linguaggio e la comunicazione” (Mucchi Faina, 2006: 22).

In questa breve panoramica del concetto di stereotipo non può essere tralasciato un fenomeno strettamente collegato agli stereotipi sociali, ma concettualmente diverso dagli stessi, ovvero il pregiudizio. I pregiudizi possono essere definiti come “atteggiamenti negativi verso persone, gruppi o altri oggetti sociali salienti, assunti a priori e mantenuti anche se riscontri empirici ne dimostrano l’infondatezza” (Palmonari et al., 2012: 275). Come gli stereotipi, anche i pregiudizi sono

categorizzazioni, comprendono cioè una dimensione cognitiva, una valutativa (nella quale vengono attribuiti valori e gerarchie) e una che genera azioni (a seconda del valore attribuito al gruppo) (*ibid*).

Come emerso dagli studi di Pettigrew e Meertens (1995) il pregiudizio si può distinguere in due categorie, il pregiudizio manifesto e quello latente o sottile. Nella società attuale il pregiudizio manifesto (ovvero il rifiuto e l'ostilità apertamente dichiarate verso un gruppo o una categoria) viene espresso meno frequentemente ed è generalmente poco accettato dalla comunità, mentre il pregiudizio latente è espresso in modi socialmente accettabili, in particolare tramite il linguaggio. Il pregiudizio sottile si esprime per esempio attraverso la difesa dei valori tradizionali e l'assenza di emozioni positive verso un gruppo o una categoria sociale.

Uno dei pregiudizi più diffusi e resistenti al cambiamento riguarda il genere femminile, basato anche sugli stereotipi di genere, che “riguardano i tratti di personalità tipici di uomini e donne, [...] spesso collegati con i ruoli maschili e femminili che ogni società attribuisce a ciascuno di noi” (Villano, 2003: 9). Sulla base di ciò le donne vengono considerate “emozionali ed espressive mentre gli uomini competenti ed indipendenti”, una visione che “rinforza e mantiene il sistema ideologico e di valori che legittima tali differenze” (Palmonari et al., 2012: 281). Questa visione inoltre rimarca le discriminazioni fra uomini e donne, poiché “socialmente si attribuisce maggior prestigio a competenza e indipendenza, considerate caratteristiche fondamentali dell'identità maschile” (*ibid.*).

È rilevante in questa sede sottolineare l'importanza della lingua non solo nel perpetuare gli stereotipi di genere, ma anche nell'evidenziare come la società, nonostante i numerosi passi avanti compiuti verso la parità di genere, manifesti ancora una predominanza maschile. La linguista Carla Bazzanella afferma infatti che “stereotipi e lingua rimandano, in modi più o meno evidenti, a una generale priorità maschile” (2010: online) e inoltre

le ideologie organizzano le rappresentazioni sociali e sottintendono dinamiche di potere che si riflettono nelle concettualizzazioni e categorizzazioni, trovando forma negli stereotipi e nella lingua. In particolare i significati lessicali sono strettamente associati a determinati schemi mentali, derivanti dall'esperienza e dalla vita sociale. (Bazzanella 2010: online).

Il rapporto fra lingua e genere, e in particolare l'applicazione di politiche volte ad un uso non sessista del linguaggio e a fornire una maggiore visibilità linguistica al genere femminile verrà approfondito nei paragrafi successivi di questo capitolo.

2.3 Lingua e genere: il caso dei femminili professionali

Come afferma Rosa Maria Bollettieri Bosinelli “il linguaggio non è uno strumento neutrale ma anzi ha un ruolo fondamentale nel plasmare la nostra percezione del mondo” (2005:47) e partendo da questo presupposto lo studio del rapporto tra lingua e genere permette di rivelare disuguaglianze e stereotipi presenti nel linguaggio, spesso così interiorizzati da essere persino difficili da individuare. In primo luogo occorre specificare che il termine genere viene utilizzato in questa sezione con diverse accezioni, una è quella di *genere sociale*, ovvero l'insieme delle

costruzioni culturali che sottostanno alla divisione sociale dei compiti e dei lavori e al conseguente consolidarsi di norme che regolano i processi di socializzazione di e tra uomini e donne a vari livelli (Laureghi e Olita, 2006: 28).

Il genere sociale è anche strettamente collegato alla realtà, ed è quindi collegato con le interazioni sociali quotidiane, nonché con gli stereotipi e le categorizzazioni che ogni parlante acquisisce ed esprime nel linguaggio di tutti i giorni. L'altra accezione che può assumere in questi paragrafi il termine genere è nel senso di *marcatatura grammaticale di genere*, presente nelle lingue che “marcano il genere con mezzi morfologici, forme pronominali, classificatori” etc. (Bazzanella, 2010: online) e in questo caso assume il significato di *genere grammaticale* che può essere femminile, maschile o, nelle lingue in cui è presente, neutro. Un esempio di come lingua e genere siano connessi è rappresentato (nelle lingue che hanno marcatatura grammaticale di genere, come per esempio le lingue romanze) dall'uso del maschile generico o ‘inclusivo’. In quest'uso i termini maschili fanno riferimento sia a uomini che donne e nello stesso contesto si inserisce anche la cosiddetta ‘servitù grammaticale’ ovvero il fenomeno linguistico per il quale, nell'accordo tra parole maschili e femminili, l'accordo avviene sempre al maschile (Bazzanella, 2010: online). Come evidenziato nelle *Linee guida per un linguaggio amministrativo e istituzionale attento alle differenze di genere* prodotte dall'Università di Padova, “il maschile generico rappresenta uno degli usi linguistici dagli effetti più discriminanti, che fa sì che solo

gli uomini siano rappresentati nella lingua, facendo rimanere invisibili le donne” (Gruppo di lavoro per un approccio di genere al linguaggio amministrativo e istituzionale, 2017: 5). Possiamo dire inoltre che questa asimmetria tra l’uso del maschile e del femminile rende evidente come “il principio del genere maschile come onnivalente e generico” (Bollettieri Bosinelli, 2005: 54) sia in realtà un modo di negare la visibilità femminile nel linguaggio, rendendolo di fatto non inclusivo. Uno degli ambiti in cui è evidente una tendenza a utilizzare il maschile sia per indicare donne che uomini è quello delle professioni, e in particolare di professioni e ruoli apicali:

chiamare le donne che fanno un certo lavoro con un sostantivo femminile non è un semplice capriccio, ma il riconoscimento della loro esistenza [...] poiché ciò che non viene nominato tende a essere meno visibile (Ghenò, 2019: 17).

La questione dei femminili professionali è più che mai attuale e alimenta un dibattito che attraversa trasversalmente la società, dalla stampa alle istituzioni fino ad arrivare ai commenti sui social network. La creazione e la diffusione dei femminili delle professioni sono un segno importante del cambiamento del ruolo della donna nella società, inoltre hanno generato (e continuano a creare) un acceso dibattito, motivi che hanno portato alla decisione di approfondire queste tematiche nell’elaborato. Le lingue cambiano e si evolvono per numerose ragioni, una di queste è sicuramente il cambiamento della società e la consapevolezza dei suoi parlanti. Si può inoltre affermare che una lingua viva viene “costruita” dai suoi parlanti, non di certo solo dalle accademie e dalle grammatiche, ed è per questo particolarmente interessante capirne l’evoluzione.

Numerosi aspetti vanno tenuti in considerazione quando si affronta il tema della formazione (o semplicemente dell’uso) delle forme femminili delle professioni. Innanzitutto la questione interessa la lingua italiana in quanto, come altre lingue romanze come per esempio il francese o lo spagnolo (di cui tratterò nelle sezioni 2.5 e 2.6), è una lingua con marcatura grammaticale del genere, che può essere femminile o maschile. In italiano tutti i sostantivi sono caratterizzati dalla loro appartenenza al genere maschile/femminile, tuttavia per i sostantivi inanimati o per concetti astratti il genere grammaticale del sostantivo non è connesso o rappresenta le caratteristiche dell’oggetto, denotando quindi arbitrarietà nell’assegnazione del genere grammaticale del sostantivo stesso (Ghenò, 2019: 52). Per quanto riguarda invece “esseri animati,

animali o persone, il genere grammaticale corrisponde per lo più al genere semantico, ossia il sesso dell'animale o dell'essere umano che la parola designa" (*ibid.*).

Per quanto riguarda la formazione del femminile, è importante soffermarsi prima su quali siano le tipologie di coppie di sostantivi maschili e femminili in italiano. Sono presenti:

i sostantivi di *genere fisso*, in cui maschile e femminile sono indicati con termini completamente diversi come per esempio padre e madre, i sostantivi di *genere promiscuo*, per i quali esiste un unico sostantivo con uno dei due generi e l'altro è formato aggiungendo maschio o femminile e di solito questo accade nel caso degli animali, come per esempio antilope maschio/femmina. Sono inoltre presenti i sostantivi di *genere comune*⁵, che hanno il maschile uguale al femminile e varia solo l'articolo, per esempio il docente e la docente. In alcuni casi, varia solo l'articolo anche al plurale, in altre i plurali maschile e femminile vengono formati con desinenze diverse. Questi sostantivi nel vocabolario riportano l'indicazione "sostantivo maschile e femminile". Troviamo infine i sostantivi di *genere mobile*, in cui il femminile e il maschile sono espressi tramite una desinenza /-o/-a, -tore/-trice, -sore/-sora ecc.) (Ghenò, 2019: 53).

Lo schema riportato espone in maniera chiara i vari modi in cui il genere femminile può essere espresso in italiano, esempio che serve come base per comprendere quale sia la maniera corretta di creare (o semplicemente declinare) i sostantivi femminili delle professioni. Il femminile delle professioni appare quindi non solo come una questione puramente grammaticale e linguistica, ma anche una culturale e sociale. Per questo motivo nelle sezioni successive verranno presentate le azioni volte a contrastare queste asimmetrie linguistiche.

2.4 La questione dei femminili professionali in Italia

Per quanto la questione dei femminili professionali e l'attenzione verso l'uso non sessista della lingua sembri estremamente recente, in Italia l'argomento viene dibattuto sin dagli anni Ottanta. Proprio nel 1987 viene pubblicato il volume *Il sessismo nella lingua italiana* a cura di Alma Sabatini per la Presidenza del Consiglio dei ministri e la Commissione Nazionale per la Parità e le Pari Opportunità tra uomo e donna, che

⁵ Insieme ai sostantivi di genere promiscuo vengo anche denominati come sostantivi *epiceni*.

viene considerato il primo documento a trattare in maniera ufficiale l'argomento. Nelle *Raccomandazioni per un uso non sessista della lingua italiana*, Sabatini elenca in maniera chiara e puntuale le forme da evitare e quelle corrette per molte professioni e cariche pubbliche, ma anche di termini che vengono spesso utilizzati al maschile per intendere in realtà sia uomini che donne. Sabatini sostiene, per esempio, che andrebbe evitata la dicitura "i diritti dell'uomo" e utilizzata invece la forma "i diritti della persona" o "i diritti umani". Tra le proposte delle linee guida ricordiamo inoltre quelle volte ad eliminare il suffisso *-essa* per molti femminili, sia perché ha acquisito un valore dispregiativo nell'uso, sia perché in passato il suffisso indicava la moglie di chi ricopriva una carica. La proposta, seppur accolta nell'uso di forme come avvocatessa e sindaca al posto di avvocatessa e sindachessa, non è stata accolta per termini come studentessa, che viene infatti correntemente utilizzato, nonostante la proposta di Sabatini di utilizzare la forma studente per entrambi i generi, modificando solo l'articolo (la/lo studente). Come affermato dalla stessa studiosa, le sue linee guida non sono un punto di arrivo, bensì l'inizio di un'analisi e segna la nascita di un dibattito che continua ancora oggi. Nelle parole dell'autrice quello che viene ricercato tramite il suo elaborato è "una riforma nel profondo dei nostri simbolismi politici, culturali, estetici, etici, che si riflettono in quella apparente superficie o parte emergente dell'iceberg che è la lingua" (Sabatini, 1987: 122).

Al pionieristico lavoro di Sabatini fanno seguito *Il Codice di stile delle comunicazioni scritte ad uso delle amministrazioni pubbliche* (1993) e successivamente il *Manuale di stile* (1997) entrambi a cura di Alfredo Fioritto. Il *Codice di stile* del '93 presenta, nel quarto capitolo, alcuni consigli e strategie rivolte alle pubbliche amministrazioni al fine di promuovere un uso non sessista e non discriminatorio della lingua. Partendo dalla concezione che in italiano il genere grammaticale maschile viene utilizzato in riferimento sia a persone di sesso maschile che di sesso femminile, ovvero con un uso non marcato, da un punto di vista sociale questo uso perpetua stereotipi sociali e alimenta un effetto di esclusione. Questo fenomeno è causato dalla concordanza tra genere grammaticale e genere naturale, che di fatto esclude le donne quando interi atti amministrativi vengono redatti utilizzando solo il maschile. Si può quindi dedurre che "l'amministrazione pubblica, attraverso i suoi atti, appare come un mondo in cui è uomo non solo chi autorizza, certifica, giudica, ma lo è anche chi denuncia, possiede immobili, dichiara ecc." (Fioritto, 1993:

49). Il capitolo si compone poi di nove punti pratici in cui vengono proposte azioni volte a

promuovere l'uso di espressioni alternative, più egualitarie, che possano incontrare il consenso della comunità dei parlanti e rispecchiare tanto i cambiamenti già radicati nelle realtà sociali tanto quelli in fieri (*ibid.*).

Appare particolarmente attuale il punto nel quale viene consigliato di volgere anche al femminile tutte le professioni o le funzioni descritte nelle offerte di lavoro e tra gli esempi notiamo architetto/architetta, una delle forme che ancora oggi, a poco più di 25 anni dalla pubblicazione, è ancora molto osteggiata.

Il 23 maggio 2007 viene poi pubblicata la Direttiva contenente le *misure per attuare parità e pari opportunità tra uomini e donne nelle amministrazioni pubbliche*, a cui fanno seguito diversi documenti che hanno l'obiettivo di promuovere la parità nella pubblica amministrazione attraverso un uso non sessista del linguaggio. Nel 2012 vengono quindi pubblicate le *Linee guida per l'uso del genere nel linguaggio amministrativo* promosse dal Comune di Firenze e redatte da Cecilia Robustelli, collaboratrice dell'Accademia della Crusca e docente presso l'Università di Modena. Nel documento vengono proposte varie strategie a beneficio della scrittura burocratica, volte ad eliminare le asimmetrie tra la forma maschile e la forma femminile e a dare visibilità al genere femminile. Numerosi sono stati i contributi in questa direzione, come per esempio l'istituzione, nel marzo 2015, del gruppo Esperti del linguaggio presso la Commissione Pari Opportunità della Presidenza del Consiglio dei Ministri e *Linee Guida per l'uso del genere nel linguaggio amministrativo del MIUR*, curate da Cecilia Robustelli e pubblicate nel 2018. Occorre altresì ricordare le numerose iniziative da parte degli Atenei volte a contrastare l'uso sessista del linguaggio, tra cui figurano per esempio le *Linee guida per un linguaggio amministrativo e istituzionale attento alle differenze di genere* promosse dall'Università di Padova nel 2017.

In seguito a questa breve panoramica storica possiamo quindi affermare che la questione del linguaggio non sessista, e in particolare l'uso al femminile delle professioni è un argomento sicuramente attuale, ma gli interventi a suo favore non sono un tema recente. Proprio su questo argomento l'Accademia della Crusca, una delle istituzioni più prestigiose che si occupa dello studio della lingua italiana, si è espressa più volte in maniera favorevole all'uso dei femminili professionali. Non solo, ma l'Accademia si è interessata in più occasioni non solo di illustrare e spiegare la

formazione di questi femminili ma anche argomentare perché il loro uso è corretto, pubblicando per esempio un intero volume dedicato alla questione, intitolato *Quasi una rivoluzione: i femminili di professioni e cariche in Italia e all'estero* a cura di Zarra e Gomez Gane. Nonostante l'uso dei femminili professionali venga quindi incoraggiato da parte dell'Accademia della Crusca e da numerosi linguisti, molte persone sono ancora restie ad utilizzarli o a declinare la propria carica o professione al femminile, specialmente perché la forma maschile di ruoli apicali è considerata di maggior prestigio sociale.

Nelle sezioni successivi verrà invece esposta la politica linguistica e il dibattito circa l'uso della forma femminile delle professioni in francese e spagnolo, due lingue che come l'italiano, esprimono il genere maschile e femminile e sono prive del genere neutro e che hanno delle istituzioni che studiano e tutelano la lingua. Per questo motivo verranno confrontate le posizioni delle Accademie (*Académie Française* e *Real Academia Española*) di Francia e Spagna. Infine il capitolo tratterà il caso della lingua inglese, oggetto di studio di questo elaborato, che, pur non presentando marcature di genere, si è dotata di strutture volte a favorire un uso non sessista della lingua.

2.5 La questione dei femminili professionali in Francia

Come già esposto nei paragrafi precedenti, la questione dei femminili professionali è molto attuale, ma non è di certo recente. Anche in Francia infatti si comincia ad analizzare la situazione dei femminili delle professioni e delle cariche a partire dagli anni 80, e più precisamente dal 1984, quando Yvette Roudy, ministra dei diritti della donna, decide di istituire una commissione per studiare la formazione del femminile nei nomi di professione. La presidente della *Commissione di terminologia per la definizione della forma femminile dei nomi di mestieri, gradi e funzioni* (la *Commission de terminologie pour la féminisation des noms de métiers, de grades et de fonctions*) tra il 1984 e il 1986 è Benoîte Groult, e nella commissione lavorano esperti ed esperte di grammatica, lingua e scrittura. L'11 marzo 1986 viene pubblicata una circolare del Primo Ministro, Laurent Fabius, contenente la guida alle regole di formazione del femminile e alcune raccomandazioni per la redazione dei documenti pubblici. Nel 1998 il Primo Ministro Lionel Jospin incarica la *Commission Générale de Terminologie* e l'*Institut National de la Langue Française* (INaLF) di studiare nuovamente la materia. L'INaLF redige quindi *Femme, j'écris ton nom*, una guida alla

formazione e all'uso del femminile di mestieri titoli o funzioni, che presenta sia una panoramica storica dell'uso dei femminili sia regole pratiche per la formazione e l'uso del femminile. INaLF afferma nella sua guida che

la parità nel lessico non è, per la maggior parte, una questione linguistica o grammaticale. Esistono già molti nomi femminili che designano occupazioni, titoli, e funzioni, gli altri si formano facilmente. La loro esistenza era latente, solo le condizioni sociali ne ritardavano l'uso. (1999: 8).

Si evince ancora una volta come la questione, anche in Francia, non sia puramente linguistica, ma porti con sé l'adeguamento della lingua al cambiamento della società e del ruolo tradizionalmente attribuito alle donne. Nonostante la guida prodotta dall'INaLF sia un grande passo avanti nell'uso e nella diffusione dei femminili professionali, presenta un limite: viene infatti consigliato l'uso del femminile solo quando ci si riferisce a una donna in quanto singola persona, mentre rimane l'uso del maschile generico quando si fa riferimento alla funzione. Nel documento si legge che

un testo di legge, decreto o ordinanza, che tratti genericamente di funzioni o titoli, sarà quindi scritto in forma maschile, di conseguenza, alcune pratiche di lingua francese che consistono nel declinare le funzioni secondo il genere non saranno seguite, fintantoché si intende la genericità della funzione. (INaLF, 1998: 38)

Questa politica alimenta l'uso del maschile generico, e inoltre non fornisce visibilità al femminile negando anche la possibilità di utilizzare la strategia dello *splitting*, ovvero l'utilizzo del sostantivo declinato al maschile e al femminile inframezzati da una barra (es. ministro/a). Tuttavia negli anni successivi le forme femminili si diffondono nella lingua e in particolare il loro utilizzo viene recepito in maniera positiva nell'ambito politico, come denota anche la sanzione pecuniaria applicata a un deputato per essersi ripetutamente rivolto a una ministra con il titolo di *Madame Le Président*, quindi non declinando il ruolo al femminile. L'*Académie Française*, l'ente francese per la tutela e il corretto utilizzo della lingua francese, prende posizione riguardo l'uso del femminile per ruoli e professioni nel 1984, nel 2002, nel 2014 e infine nel 2019. L'*Académie*, infatti, si pronuncia in maniera contraria all'uso dei femminili per i ruoli e gli incarichi fin dal 1984, sostenendo che la questione dei femminili professionali non possa essere banalizzata riducendola a una semplice opposizione del termine femminile e maschile poiché gli usi del genere grammaticale sono in realtà una rete molto più complessa di come appare, e che la designazione del

sempre gioca solo un “ruolo minore” nell’assegnazione del significato (1984). L’intervento sostiene inoltre che i cambiamenti deliberati (sottintendendo forse anche un’arbitrarietà di queste scelte) possono creare confusione e disordine, ed è quindi consigliabile non alterare il sottile equilibrio della lingua (*ibid.*). Nel 2002 l’*Académie* continua a sostenere la sua posizione, affermando che

come già sottolineato nel 1984, la progressiva instaurazione di una reale parità tra uomini e donne nella vita politica ed economica rende indispensabile la conservazione di denominazioni collettive e neutrali, e quindi il mantenimento del genere non marcato, ove possibile. Al contrario, la scelta sistematica e sconsiderata delle forme femminili stabilisce, all'interno del linguaggio stesso, una segregazione controproducente. (2002: online).

La posizione dell’*Académie* è ancora una volta quella di differenziare fra l’uso del femminile per i mestieri e l’uso del femminile per i ruoli: nessuna opposizione all’uso del femminile per mestieri e professioni, ma l’istituto non rileva nessuna esigenza di volgere al femminile i nomi dei ruoli e degli incarichi, dato che in quanto ruoli sono generici e non riferiti alla singola persona. Il punto di vista è ancora una volta a favore del maschile generico, considerato “inclusivo” e di conseguenza utilizzabile per designare un ruolo ricoperto da un uomo o da una donna. Nonostante le resistenze della più prestigiosa istituzione della lingua francese, negli anni i femminili dei ruoli e delle cariche si sono diffusi sempre di più nel linguaggio politico e giornalistico, portando l’*Académie Française*, nel 2019, a dover ritornare sulla questione. Viene quindi pubblicato un nuovo rapporto sulla creazione del femminile dei sostantivi che designano incarichi o ruoli istituzionali, che per la prima volta registra e ammette l’uso di queste forme, anche se non lo incoraggia. Il rapporto si limita infatti a fornire delle regole grammaticali sulla formazione del femminile, sostenendo che sia comunque possibile utilizzare la forma maschile per designare incarichi ricoperti da donne. Dalle posizioni dell’istituto emerge chiaramente come la questione non sia puramente linguistica, ma che ci siano profonde radici sociali nella volontà di non utilizzare il femminile nelle professioni, e in particolare nei ruoli che tradizionalmente sono associati a un maggior prestigio sociale, come incarichi pubblici o politici. Il punto di vista finale dell’*Académie* è il seguente:

se la facilità con cui si sono affermate le forme femminili che designano le funzioni ai vertici della gerarchia permette di fare il punto sui cambiamenti d'uso degli ultimi

vent'anni, resta il fatto che, dato che alcune donne che ricoprono incarichi, incarichi che ancora oggi sono spesso occupati da uomini, hanno espresso una preferenza per essere designate nelle loro funzioni al maschile, non c'è motivo per cui questo uso dovrebbe essere modificato. (2019: 20).

Il caso francese può essere considerato opposto a quello italiano, dove una grande attenzione istituzionale e degli enti linguistici non corrisponde a un'applicazione e diffusione capillare dei femminili di ruoli e professioni. La situazione in Francia è invece ribaltata: sono stati infatti proprio i parlanti e in particolare i termini utilizzati nel linguaggio politico e giornalistico a "spingere" l'*Académie Française* a dover accettare e includere queste forme nei dizionari e nelle grammatiche.

2.6 La questione dei femminili professionali in Spagna

Come in Italia e in Francia, anche in Spagna la questione del femminile per ruoli e professioni si impone progressivamente a partire dagli anni Ottanta, tuttavia la questione non è stata affrontata in prima battuta dal governo centrale, delegando invece l'argomento ai poteri periferici come ministeri, comunità autonome, associazioni o università. Proprio per questo motivo le guide sull'uso non sessista del linguaggio sono numerosissime e non sono presenti solo per la lingua spagnola, ma sono state elaborate anche per il catalano, e successivamente, per il basco e il gallego. Tra le prime guide degli anni Ottanta è particolarmente rilevante la guida prodotta dal *Ministerio de Educación y Ciencias* (Ministero dell'Educazione e delle Scienze) nel 1988, che tra i suoi obiettivi ha quello di eliminare gli stereotipi di genere nei testi scolastici, oltre a proporre di declinare al femminile i sostantivi di ruoli e professioni esercitati da una donna:

evitare l'uso esclusivo del maschile per nomi di professioni, mestieri e carriere, che indicano posizioni di prestigio, quando il femminile esiste e spesso viene utilizzato solo per lavori gerarchicamente inferiori e tradizionalmente legati al "ruolo femminile". (1988:20).

L'argomento viene poi ripreso e ampliato nel volume di Eulàlia Lleidò Cunill, *Las profesiones de la A a la Z*, edito dall'*Instituto de la Mujer* nel 1996 (ristampato in edizione aggiornata nel 2006) che contiene non solo le regole di formazione del femminile, ma anche una lista esaustiva di professioni declinate al maschile e al femminile. Il *Ministerio de Sanidad, Servicios sociales e Igualdad* nel 2015 ha censito le guide all'uso non sessista del linguaggio pubblicate da enti locali, comunità

autonome, università e associazioni classificandole tematicamente in università, amministrazione, comunicazione e pubblicità, sport e cultura, disabilità, istruzione, occupazione e rapporti di lavoro, sensibilizzazione generale e risorse web, ambito legale, ambito medico, società civile o scienza, tecnologia e infine ambiente; annoverando più di 100 guide nelle varie tematiche. Il grande interesse per questo argomento non è limitato solo alle istituzioni, difatti anche la *Real Academia Española* (RAE), istituzione responsabile dell'uso corretto della lingua spagnola e della pubblicazione del dizionario normativo della lingua spagnola, ha preso più volte posizione sia riguardo al linguaggio non sessista che all'uso del femminile per professioni e ruoli. Nel 2012, infatti, Ignacio Bosque ha pubblicato per la RAE una dettagliata critica alle proposte contenute in numerose linee guida sull'uso non sessista del linguaggio, documento sottoscritto da 26 membri della *Real Academia*. Il documento redatto da Bosque si concentra principalmente sul ribadire la validità del maschile generico in tutti i contesti, rimarcando l'inutilità e anzi la dannosità che proposte come lo sdoppiamento (es. lavoratori e lavoratrici) o l'utilizzo di termini generici (l'infanzia al posto di bambini e bambine) portano. Per quanto riguarda invece l'uso del femminile di cariche o professioni, Bosque afferma:

nessuno considera controversa la proposta di estendere la formazione di coppie morfologiche ai nomi di professioni e ruoli, oggi universalmente accettata. Tuttavia queste linee guida non sembrano ammettere che una professionista che lavora nella magistratura possa scegliere tra l'essere definita *juez* (*il giudice*) o *la jueza* (*la giudice*) [...], nonostante in molti casi sia stato riscontrato che esistono preferenze geografiche, oltre a quelle personali, per l'una o l'altra denominazione. (Bosque 2012: 8, traduzione e corsivo miei).

In un'intervista rilasciata a *BBC Mundo* pubblicata l'8 marzo 2012, la filologa dell'Università di Alcalá de Henares Mercedes Bengoechea spiega che le stesse donne spesso si vedono attraverso "occhi maschili" e che fin da piccole le bambine imparano che quando si dice bambini a volte saranno comprese anche loro, altre volte no. In questo modo "entrano in un processo di autocensura, nel quale in alcune occasioni si assentano dalla realtà perché la lingua non le rappresenta" (Bengoechea, 2012: online). All'interno del dibattito sembra quindi esserci ancora una preponderanza dell'uso del maschile generico per indicare sia uomini che donne anche nelle professioni, tuttavia si riscontra anche un'apertura maggiore verso la possibilità di declinare al femminile i sostantivi che denotano professioni e ruoli. Come esposto dalla filologa della

Fundación del Español Urgente (Fundéu BBVA) Judith González Ferrán, alcune forme femminili che suscitano tanto clamore erano già presenti nella lingua spagnola, ciò che è cambiato è il loro uso, che si è adeguato al cambiamento della società. L'esempio presentato da González Ferrán è il caso di *presidenta*, presente nei dizionari fin dal 1803, che oggi ha come prima accezione “chi presiede” e non “moglie del presidente” come veniva invece definita nelle precedenti edizioni dei dizionari (González Ferrán, 2017: online). González Ferrán inoltre afferma che la *Fundación del Español Urgente* si pone come obiettivo quello di fornire una forma femminile valida e corretta dal punto di vista grammaticale, per permettere a tutti di utilizzarli nella lingua e, di conseguenza, standardizzare questi termini permettendone un uso sempre più diffuso (*ibid.*). In ultimo si riporta il punto di vista sulla questione del maschile generico, sul quale *Fundéu BBVA* sostiene che utilizzare forme maschili delle professioni per “riferirsi alle donne configura nella mente del destinatario un panorama non sempre corrispondente alla realtà” (*ibid.*).

Possiamo quindi notare da questo confronto come spesso le “Accademie” siano più conservatrici della società, e siano di conseguenza più restie al cambiamento naturale della lingua. Se da una parte l'obiettivo di “conservare” l'integrità e la norma linguistica è apprezzabile, in contesti come quello esposto nelle sezioni precedenti si può notare come invece questo sia un freno alla volontà di rendere la lingua più attenta alle questioni di genere e più inclusiva.

2.6 La questione nella lingua inglese

L'inglese è una lingua a bassa marcatura di genere che, a differenza delle lingue presentate nei paragrafi precedenti, potrebbe considerarsi meno rilevante per uno studio su queste tematiche, tuttavia anche lingue di questo tipo non sono esenti dalle questioni legate al sessismo linguistico, dimostrando così che il tema non è puramente linguistico, ma che ha profonde radici sociali. Fin dalla fine degli anni Settanta sono emersi lavori riguardanti questo tema come *Guidelines for non-sexist language in APA journals* del 1978 o *The Myth of the Neutral 'Man'* di Janice Moulton pubblicato nel 1981, dove si evidenzia come siano presenti disparità di genere per quanto riguarda titoli professionali, nomi collettivi e stato civile. Secondo Kate Miller e Casey Swift “terms used of a class or group that are not applicable to all its members are false generics” (1988: 11); per questa ragione il prefisso (o il suffisso) *man* è considerato un

falso generico, di conseguenza anche l'uso del maschile generico risulta non inclusivo. Ai termini *mankind* o *manpower* usati per indicare in generale l'umanità o il personale andrebbero preferiti i termini che possono essere considerati a tutti gli effetti inclusivi, come per esempio *humankind* o *human resources*.

Le linee guide e le proposte avanzate sono state accettate positivamente soprattutto nel campo delle professioni, dove alcune occupazioni contenenti il suffisso *-man*, come nel caso di *chairman*, *policeman*, o *fireman*, sono state sostituite con i termini *gender-neutral* (ovvero che non si riferiscono né al genere femminile né al genere maschile) *chair/chairperson*, *police officer* e *firefighter*. In un articolo pubblicato sul giornale *The Guardian* nel 2016 la professoressa di linguistica applicata dell'Aston University, Judith Baxter, sostiene che

there is a strong view among applied linguists that while words can't change attitudes overnight, the continuous, repetitive use of gender-neutral terms does help people to change their view of reality over time. If we use non-gendered words most of the time, we begin to see people and professions as non-gendered too. (2016: online)

Sono numerose le politiche linguistiche messe in atto per riconoscere ed eliminare gli stereotipi di genere e l'uso sessista della lingua inglese, e in questo paragrafo sono state nominate solo alcune tra le più comuni e diffuse. Il ruolo del movimento femminista nell'individuazione e nell'attuazione di queste politiche è stato fondamentale, tanto nella lingua inglese come nelle altre lingue esposte in precedenza.

2.7 Conclusione e nuovi spunti

In conclusione di questo capitolo vorrei soffermarmi su alcuni aspetti che questa panoramica può offrire riguardo al rapporto tra genere e lingua. Possiamo affermare che la resistenza al cambiamento, specialmente quando si tratta di termini che designano ruoli e professioni apicali, è frutto di una cultura maschilista. Cambiare questi termini rende evidente il nuovo ruolo che le donne hanno assunto nella società, mettendo in luce i risultati che anni di lotte hanno portato ad ottenere. Occorre quindi diffondere la consapevolezza di questi temi tra i parlanti, in particolare tra coloro che “vivono” di lingua, come i traduttori. Non solo, ma la diffusione di questi termini dovrebbe aumentare, considerando che la presenza femminile si attesta in crescita anche in alcune carriere che sono state per lungo tempo appannaggio solo maschile. In questo senso è emblematico il caso della magistratura italiana, dove i dati forniti dal

Consiglio Superiore della Magistratura per il 2018⁶ rivelano un numero maggiore di donne nella professione: sono infatti 5013 le magistrate e 4388 i magistrati, nonostante le donne abbiano avuto accesso alla magistratura solo nel 1965. È inoltre fondamentale diffondere sempre di più l'uso di queste forme e di tutte le strategie volte a rendere le lingue più rispettose delle differenze di genere nell'ottica dell'intelligenza artificiale, dato che sempre più applicazioni utilizzano corpora e testi che contengono stereotipi e bias, come già esposto nel precedente capitolo. Occorre dare visibilità al ruolo della donna e alle lotte che sono state compiute e che vengono ancora oggi portate avanti al fine di ottenere l'uguaglianza di genere. Un esempio di come si potrebbero modificare le abitudini dei parlanti è quello di iniziare a utilizzare quello che potrebbe definirsi “femminile inclusivo” o, in un'altra ottica, un accordo rispetto al numero maggiore di entità maschili e femminili a cui si sta facendo riferimento. Un esempio calzante di questa politica è per esempio il sito del Dipartimento di Traduzione e Interpretazione (DIT) di Forlì, un Dipartimento con una forte prevalenza di donne nel personale docente e amministrativo. Nel sito del Dipartimento⁷ tutte le professioni sono state declinate solo al femminile, allegando una nota che spiega le motivazioni di questa decisione:

Al DIT non ci sono anche ricercatori, assegnisti, professori? Ovviamente la risposta è sì. Ma in questa pagina abbiamo deciso di utilizzare, quando la lingua italiana imponeva di scegliere, forme al femminile. Questa decisione evita soluzioni che sarebbero andate a discapito della leggibilità e accessibilità del sito (asterisco, alternanza maschile/femminile) e allo stesso tempo dà conto del numero elevato di donne che lavorano nella struttura e dell'attenzione che il Dipartimento ha per la comunicazione e i rapporti che legano la lingua alle tematiche di genere). (2020: online)

È anche da questi gesti che si costruisce una lingua, e forse quello sopra riportato è un primo passo verso un nuovo cambiamento nell'uso della lingua.

⁶ Donne in magistratura: <http://bit.ly/2ZD6cSZ>

⁷ <https://dit.unibo.it/it/dipartimento/persone>

Capitolo III: Metodo

3.1 Studi precedenti

Di seguito verranno esposte alcune strutture di frasi che sono state utilizzate negli studi precedenti nell'ambito dell'analisi degli stereotipi di genere nella traduzione automatica. Successivamente, invece, si illustra la struttura che verrà impiegata in questo studio di caso, ovvero una struttura frasale *ad hoc*.

Gli studi precedenti a questo elaborato che hanno utilizzato set di frasi per l'individuazione di stereotipi di genere e inesattezze linguistiche nella TA, attualmente sono quattro:

1. Stanovsky et al. (2019) nel quale vengono utilizzati i set Winogender e WinoBias creando un set bilanciato denominato WinoMT, basati sugli schemi di Winograd.
2. Prates et al. (2018) nel quale utilizza una struttura frasale molto semplice, che nella traduzione viene resa con “<he/she> is a <profession>”.
3. Monti (2017) propone di studiare alcuni fenomeni linguistici che mostrano in maniera evidente gli stereotipi di genere e/o i problemi legati al riconoscimento del femminile da parte dei sistemi di TA, utilizzando vari esempi che verranno riportati interamente.
4. Escudé Font e Costa-jussà (2019) propongono una frase *ad hoc* per studiare il fenomeno dei *gender bias* nella TA, costruendo una struttura frasale ripetibile e con diversi elementi che possono essere modificati nello studio. La struttura frasale elaborata è “I’ve known <him/her/proper noun> for a long time, my friend works as a/an <occupation>”.

Qui di seguito verranno illustrati gli studi elencati, soffermandosi in particolare sulle motivazioni che mi hanno portato ad escludere o a prendere spunto dagli studi precedenti per elaborare la struttura frasale proposta in questo studio di caso. È opportuno specificare che in questa sezione viene esposta la metodologia impiegata nelle quattro ricerche al fine di valutarne il possibile impiego nella mia proposta di studio di caso. Partendo dallo studio di Stanovsky et al. presentato al punto 40 più sopra, si espongono le motivazioni che hanno portato a non riprodurre la struttura frasale elaborata in questo studio nella mia proposta. Occorre innanzitutto ricordare

che è uno studio molto esteso (il set contiene infatti più di 3000 frasi), con più partecipanti (professioni) in una stessa frase, rendendo di fatto complicata una valutazione manuale e accurata, come si intende invece effettuare in questo elaborato. Lo studio di Stanovsky, basato sui precedenti set proposti da Rudinger et al. (2018) e Zhao et al. (2018), già illustrati in §1.3.6.1, utilizza frasi basate sullo schema di Winograd. Le frasi di questo schema, seppure molto efficaci per valutare l'accuratezza dei sistemi di TA riguardo la coreferenza, sono ambigue sia in italiano che in inglese, rendendo impossibile verificare la correttezza della traduzione. Per meglio comprendere le obiezioni riportate, riporto di seguito un esempio di frasi utilizzate da Stanovsky e al.⁸:

(1) *The farmer* did not want to talk with the writer because *she* was busy with irrigation.

(2) The farmer did not want to talk to *the writer* because *he* was burying himself in writing a new novel.

Le frasi sopra citate, seppure molto interessanti dal punto di vista degli stereotipi, permettono di identificare il genere solo di uno dei due partecipanti dell'azione, lasciando inferire al sistema (ma anche a un qualsiasi parlante di italiano che abbia intenzione di tradurre queste frasi) il genere dell'altro partecipante all'azione. Nella prima frase (41) possiamo dedurre che il sostantivo “farmer” vada tradotto al femminile grazie al pronome “she”, tuttavia non sappiamo se il sostantivo “writer” vada tradotto al maschile o al femminile. Una struttura frasale di questo tipo potrebbe essere utile per capire gli stereotipi di genere, per esempio indagando su quali professioni vengano tradotte sempre al maschile o al femminile in assenza di un pronome che li associ a un uomo o a una donna, tuttavia non permette né di osservare la traduzione dal punto di vista del genere di entrambi i termini riferiti a una professione, né permette una valutazione qualitativa della traduzione, poiché le frasi sono molto diverse tra di loro dal punto di vista semantico. Infine un ulteriore motivo che mi ha portato ad escludere l'ipotesi di utilizzare questo set di frasi è che nello studio originale si intende testare, oltre alla presenza degli stereotipi di genere, anche la capacità da parte dei sistemi di TA di identificare correttamente l'antecedente sulla base della semantica della frase. Per le finalità del mio studio, cioè quelle di verificare

⁸ Disponibili all'indirizzo https://github.com/gabrielStanovsky/mt_gender

l'eventuale presenza di stereotipi di genere nella TA, non è necessario inserire questa variabile che comunque non verrebbe verificata e questo aspetto limiterebbe lo studio di altri fenomeni più rilevanti.

Lo studio di Prates et al. et al. (2018) presenta un limite che potremmo definire “linguistico”, infatti le frasi originali di questo studio sono state elaborate in 12 lingue che non hanno marcatura di genere, e di conseguenza non hanno pronomi maschili o femminili. Le frasi sono state tradotte da queste lingue verso l'inglese, mentre nel mio caso la direzione della traduzione è dall'inglese verso l'italiano. La struttura frasale utilizzata da Prates et al. et al. (“<he/she> is a <profession>”) inoltre non permette di studiare né la coreferenza, né altri aspetti linguistici che possono far emergere una (possibile) maggiore propensione per l'uso del maschile in determinate professioni o contesti, come per esempio nell'accordo tra il gruppo nominale e il soggetto/referente della frase o tra aggettivo e soggetto/agente della frase.

Lo studio di Monti (2017) si differenzia dagli altri che vengono presentati in queste pagine poiché è il primo studio che si focalizza solo sulla combinazione inglese italiano e pone l'accento su questioni linguistiche, come già esposto in §1.3.6.3. Nel suo elaborato, Monti traduce automaticamente frasi che riguardano:

- Accordo soggetto – predicato nominale
- Accordo soggetto|oggetto – complemento predicativo del soggetto|dell'oggetto
- Accordo nome – apposizione
- Accordo nome – riferimento anaforico/cataforico
- Accordo nome - participio passato
- Genere sociale

Per ogni aspetto vengono proposte alcune frasi-esempio tradotte da diversi sistemi di TA (neurale, statistico e basato su regole). Verranno qui di seguito riportati gli esempi utilizzati per ogni aspetto e infine alcune considerazioni sulle frasi utilizzate.

Accordo soggetto - predicato nominale

- (1) Mary is a diligent employee.
- (2) Mark is a diligent nurse.
- (3) Mark is an efficient nurse.

(4) The speaker is my sister.

Accordo soggetto/oggetto - complemento predicativo del soggetto/complemento predicativo dell'oggetto

(7) Jane seems very nervous.

(8) My sister was nominated sole director.

(9) The board appointed my sister as sole director.

Accordo nome – apposizione

(10) Mary, our doctor, went away few minutes ago.

(11) Our doctor, Mary, went away few minutes ago.

Accordo nome - participio passato

(12) Mary is happily married.

(13) Though happily married to a Jewish man, Virginia Woolf wrote of Jewish characters in stereotypical archetypes.

Accordo nome - riferimento anaforico/cataforico

(14) The student studied really hard for her test.

(15) If you want it, there is a salad in the fridge.

Genere sociale

(16) John is a secretary.

(17) John is a secretary of state. (Monti 2017: 424-430)

Le frasi utilizzate da Monti offrono spunti interessanti di riflessione, tuttavia non possono essere considerate una base per costruire un set di frasi riproducibili. La separazione dei fenomeni osservati, seppure utile a chiarire la tematica trattata, non permette a mio avviso di confrontare i dati che si potrebbero ricavare da uno studio di questo tipo, dato che le frasi utilizzate sono diverse per ogni caratteristica (e anche all'interno della stessa). Ritengo che non sia quindi possibile confrontare le frasi fra di loro, e utilizzare una struttura di questo tipo produrrebbe, a mio avviso, una frammentazione dei risultati, che impedirebbe una valutazione complessiva delle frasi.

L'ultimo studio presentato è quello di Escudé Font e Costa-jussà (2019) nella combinazione linguistica inglese-spagnolo. I due studiosi hanno creato un sistema di TA e hanno elaborato una struttura frasale per verificare la presenza di stereotipi di genere all'interno del sistema da loro creato. La struttura frasale elaborata è "I've

known <him/her/proper noun> for a long time, my friend works as a/an <occupation>”. Questa permette di verificare sia come viene tradotto il termine “friend” (ambiguo dal punto di vista del genere dato che può essere tradotto sia al maschile che al femminile), sia di verificare come vengono tradotte la professione e il gruppo nominale (*my friend*) sulla base del coreferente (*him/her/proper noun*). La struttura frasale citata, permette di verificare alcuni fenomeni, ma presenta alcuni limiti in un’ottica di traduzione; ad esempio nel caso di professioni tradotte con termini epiceni/ambigenere (es. dentista) per verificare se la professione si riferisce ad un uomo o a una donna occorre controllare l’accordo con il gruppo nominale. Allo stesso modo nella traduzione non è presente l’articolo prima della professione, rendendo poco chiaro o non verificabile il genere della professione in alcuni casi (es. sostantivi epiceni). Per questi motivi ho scelto di non utilizzare pienamente la frase proposta da Escudé Font e Costa-jussà e nella sezione successiva verrà esposta la struttura frasale elaborata.

3.2 Struttura frasale

In seguito all’analisi delle frasi proposte dagli studi precedenti, è stato deciso di creare un set frasale *ad hoc* includendo un coreferente della professione al fine di rendere esplicito il rapporto tra genere del pronome e genere della professione. Nella struttura frasale elaborata è inoltre presente un aggettivo dimostrativo prima della professione in modo da permettere di identificare chiaramente il genere maschile o femminile anche nelle professioni tradotte con termini epiceni. La struttura frasale prodotta è quindi: “I’ve known <him/her/proper noun> for a long time, this <occupation> is one of the best in <his/her> field”, e un esempio della stessa è “I’ve known *him* for a long time, this *nurse* is one of the best in *his* field” nel caso del coreferente maschile e “I’ve known *her* for a long time, this *nurse* is one of the best in *her* field” nel caso del coreferente femminile. Le professioni utilizzate nel set sono estratte dal *Bureau of Labor Force Statistics* degli Stati Uniti; nella sezione successiva verranno spiegati nel dettaglio i criteri di selezione e il campione di professioni scelte per lo studio.

3.3 Selezione delle professioni

Una parte fondamentale della costruzione dello studio è rappresentata dalla scelta delle professioni da utilizzare nel set frasale elaborato. Come in studi precedenti (Prates et

al. 2018, Escudé Font e Costa-jussà 2019) sono stati utilizzati i dati prodotti dallo *US Bureau of Labor Force Statistics*, nello specifico sono stati considerati i dati sulle professioni ottenuti dal *Current Population Survey* (Censimento della popolazione) del 2019. Le statistiche utilizzate mostrano il numero di lavoratori totali per ogni professione (con un totale di più di 500 occupazioni suddivise in 22 macrocategorie). Qualora la professione superi le 50.000 unità, viene riportata anche la percentuale di donne presenti nella stessa. La selezione è stata effettuata manualmente, e sono stati applicati alcuni criteri di esclusione. Sono state in primis considerate solo le professioni con più di 50.000 lavoratori, in modo da disporre della percentuale di donne nelle professioni, successivamente sono stati considerati solo i termini che potessero riferirsi sia a uomini che donne, escludendo, per esempio, termini come *actor/actress* e *waiter/waitress*. Sono poi state escluse le occupazioni culturalmente specifiche, ovvero quelle che pur presentando una possibile traduzione in italiano, non hanno un equivalente diretto e univoco in Italia, come per esempio *practitioner nurse* o *millwright*. Sono state inoltre scartate le professioni con etichette troppo generiche, come per esempio *other education, training, and library workers* o *computer occupations*.

Per estrapolare il campione finale delle professioni utilizzate nello studio sono poi state selezionate un numero uguale di professioni a maggioranza femminile e a maggioranza maschile, considerando una professione a maggioranza femminile o maschile quando il numero relativo è superiore al 50% del totale. Il campione finale è di 56 professioni, composto da 28 professioni a maggioranza maschile e 28 a maggioranza femminile. In realtà è bene ricordare che molte professioni non sono fortemente polarizzate, mostrando quindi una vicinanza tra le percentuali di lavoratori e lavoratrici. Inoltre ognuna delle 22 categorie professionali definite dallo *US Bureau of Labor Force Statistics* è rappresentata da almeno una professione scelta per il campione, garantendo così una rappresentatività generale dello stesso. In aggiunta, alcune etichette utilizzate dal *Bureau* contengono al loro interno più di una professione, e, qualora selezionate, i termini sono stati considerati separatamente. Occorre però precisare che presentano le stesse percentuali di occupazione femminile, come per esempio nel caso di *Judges, magistrates, and other judicial workers*. In questo caso, l'etichetta *other judicial workers* è stata scartata perché troppo generica, mentre i termini *judge* e *magistrate* sono stati separati. La tabella 1 riporta le professioni selezionate ordinate secondo la

percentuale di lavoratrici nella stessa; ove necessario sono riportate tra parentesi delle indicazioni specifiche al fine di chiarire l'ambito dell'occupazione.

Occupazione	Percentuale di donne
Electrician	2,2
Carpenter	2,8
Firefighter	3,3
Pilot	7,5
Painter (construction)	8,9
Engineer (civil)	13,9
Taxi driver	16,8
Police officer	17,6
Software developer	18,7
Programmer	20,3
Chef	22
Architect	24,5
Butcher	26,9
Chief executive	27,6
Dentist	33,9
Lawyer	36,4
Mail carrier	36,6
Singer	36,7
Janitor	37,2
Physician	40,8
Surgeon	40,8
Cook	41,7
Director	42,5
Tour guide	43,6
Bus driver	45,3
Professor (post-secondary teacher)	47,4
Scientist (biological)	47,7
Photographer	49,3
Judge	52,5
Magistrate	52,5
Bartender	53,1
Artist	54
Writer (technical)	57,8
Baker	60,4
Pharmacist	60,4
Accountant	61,7
Appraiser (claim)	62,1
Veterinarian	68,3
Tax collector	69,2

Health worker	70,8
Cashier	71,2
Model	73,3
Tailor	75,4
Psychologist	79,7
Librarian	79,9
Teacher (elementary and middle school)	80,5
Flight attendant	81,3
Social worker	81,9
Teller	84,7
Occupational therapist	88
Nurse	88,9
Cleaner	89
Nutritionist	92,1
Hairdresser	92,3
Speech therapist	95,8
Dental hygienist	96

Tabella 1: professioni selezionate e relativa percentuale di occupazione femminile (fonte: US Bureau of Labor Statistics, <https://www.bls.gov/cps/cpsaat11.htm>)

Le professioni selezionate possono inoltre essere divise in cinque intervalli a seconda della numerosità di donne nella professione. Nella tabella 2 sono riportati gli intervalli delle percentuali e il numero di professioni che appartengono a quella percentuale nel campione di professioni scelte per lo studio. L'intervallo fra il 40 e il 60%, inoltre, è ulteriormente suddiviso per mostrare il numero delle professioni al di sopra e al di sotto della soglia del 50% che ho considerato come soglia per definire una professione a maggioranza maschile o femminile. Il campione ottenuto è quindi bilanciato non solo fra professioni con una percentuale maggiore o minore del 50% di donne nell'occupazione, ma anche fra gli intervalli presi in esame, rendendolo quindi rappresentativo di tutti gli intervalli analizzati.

Intervallo	Numero professioni	
0-20%	10	
20-40%	11	
40-60%	14	7
		7
60-80%	11	
80-100%	10	

Tabella 2: intervalli percentuali e numero di professioni selezionate

Nelle tabelle sotto riportate (tabelle 3-7) gli intervalli prima esposti sono stati rinominati a seconda della percentuale di donne della professione e conseguentemente con la prevalenza maschile o femminile delle occupazioni all'interno dell'intervallo.

Intervallo	Professioni
0-20%: primo intervallo, forte prevalenza maschile (10)	Carpenter Firefighter Pilot Painter (construction) Engineer (civil) Taxi driver Police officer Software developer Programmer

Tabella 3: secondo intervallo e professioni

Intervallo	Professioni
20-40%: secondo intervallo, prevalenza maschile (11)	Chef Architect Butcher Chief executive Dentist Lawyer Mail carrier Singer Janitor Physician Surgeon

Tabella 4: secondo intervallo e professioni

Intervallo	Professioni
40-60%: terzo intervallo, lieve prevalenza maschile o femminile (14)	Cook Director Tour guide Bus driver Professor (post-secondary teacher) Scientist (biological) Photographer Judge Magistrate bartender artist Writer (technical) Baker Pharmacist

Tabella 5: terzo intervallo e professioni

Intervallo	Professioni
60-80%: quarto intervallo, prevalenza femminile (11)	Accountant Appraiser (claim) Veterinarian Tax collector Health worker cashier Model Tailor Psychologist Librarian Teacher (elementary and middle school)

Tabella 6: quarto intervallo e professioni

Intervallo	Professioni
80-100%: quinto intervallo, forte prevalenza femminile (10)	Flight attendant Social worker Teller Occupational therapist Nurse cleaner Nutritionist Hairdresser Speech therapist Dental hygienist

Tabella 7: quinto intervallo e professioni

3.4 Confronto con dati italiani

Per una corretta analisi sono stati ricercati anche i dati relativi alle percentuali di donne nelle professioni in Italia, ipotizzando la possibilità che le percentuali di presenza femminile potessero variare tra i dati statunitensi e quelli italiani. I dati Istat non sempre definivano il sesso degli occupati nelle professioni, per questo ho integrato con i dati ricavati sia dagli albi professionali che dalle associazioni delle singole professioni. Nelle note sono fornite tutte le fonti, i dati utilizzati sono i più recenti che è stato possibile reperire, generalmente sono stati considerati attuali i dati dal 2015 al 2019. Dove non è stato possibile reperire dati è stata inserita una barra.

Occupazione (EN)	Percentuale di donne (USA)	Percentuale di donne (Italia)
Electrician	2,2	>1
Carpenter	2,8	>1
Firefighter	3,3	2,13 ⁹
Pilot	7,5	16
Painter (construction)	8,9	2
Engineer (civil)	13,9	15,3
Taxi driver	16,8	4
Police officer	17,6	15,4 ¹⁰
Software developer	18,7	19
Programmer	20,3	11
Chef	22	30
Architect	24,5	/

⁹ <https://www.interno.gov.it/it/sala-stampa/dati-e-statistiche/presenza-donne-nel-ministero-dellinterno>

¹⁰ Ibid.

Butcher	26,9	17
Chief executive	27,6	15
Dentist	33,9	19,6 ¹¹
Lawyer	36,4	47,7 ¹²
Mail carrier	36,6	38
Singer	36,7	21
Janitor	37,2	26
Physician	40,8	43,6 ¹³
Surgeon	40,8	43,6
Cook	41,7	72
Director	42,5	/
Tour guide	43,6	63
Bus driver	45,3	3
Professor (post-secondary teacher)	47,4	40,2 ¹⁴
Scientist (biological)	47,7	/
Photographer	49,3	26
Judge	52,5	53,3 ¹⁵
Magistrate	52,5	53,3
Bartender	53,1	59
Artist	54	46
Writer (technical)	57,8	47
Baker	60,4	24
Pharmacist	60,4	66 ¹⁶
Accountant	61,7	66
Appraiser (claim)	62,1	35
Veterinarian	68,3	46,5 ¹⁷
Tax collector	69,2	57
Health worker	70,8	/
Cashier	71,2	83
Model	73,3	76
Tailor	75,4	83
Psychologist	79,7	83 ¹⁸
Librarian	79,9	74
Teacher (elementary and middle school)	80,5	89,3 ¹⁹
Flight attendant	81,3	74

¹¹ <https://portale.fnomceo.it/8-marzo-la-medicina-e-donna-lodontoiaatria-tiene-il-passo/>

¹² www.cassaforense.it/media/7194/rapporto-censis-2018.pdf

¹³ <https://portale.fnomceo.it/8-marzo-la-medicina-e-donna-lodontoiaatria-tiene-il-passo/>

¹⁴ <http://ustat.miur.it/documenti/2019-notiziario-statistico-n-2/>

¹⁵ <http://bit.ly/2ZD6cSZ>

¹⁶ <https://www.federfarma.it/Farmaci-e-farmacie/Notizie-e-dati-dall-Europa/Numero-di-farmacisti.aspx>

¹⁷ <http://www.fnovi.it/node/47988>

¹⁸ <https://www.psy.it/dati-statistici>

¹⁹ <https://dati.istruzione.it/opendata/opendata/catalogo/elements1/?area=Personale%20Scuola>

Social worker	81,9	/
Teller	84,7	53
Occupational therapist	88	/
Nurse	88,9	77,3 ²⁰
Cleaner	89	74
Nutritionist	92,1	/
Hairdresser	92,3	66
Speech therapist	95,8	/
Dental hygienist	96	/

Tabella 8: professioni selezionate e relativa percentuale di occupazione femminile negli Stati Uniti e in Italia

3.5 Analisi corpus-based

Successivamente alla preparazione delle frasi e alla selezione delle professioni le frasi, queste sono state tradotte dai due sistemi di traduzione automatica selezionati per lo studio, ovvero Google Translate e DeepL. Al fine di analizzare nel dettaglio alcuni sostantivi riferiti alle professioni prese in esame e il loro uso nella lingua, è stata effettuata un'analisi degli stessi con l'utilizzo di corpora. In particolare, è stato utilizzato il corpus itTenTen16, un corpus di testi italiani raccolti dal web tra la fine di maggio e la metà di agosto 2016; parte della raccolta di corpora chiamati "TenTen", definiti dai ricercatori che li hanno sviluppati come "a set of the web corpora built using the same method with a target size 10+ billion words²¹". Il corpus itTenTen è costituito da 4,9 miliardi di parole, ed è stato taggato utilizzando il tagset elaborato da Marco Baroni. Il corpus è stato interrogato tramite la piattaforma Sketch Engine²² e tutte le informazioni relative alle funzionalità utilizzate sono state reperite dal sito web dello stesso. In particolare sono state utilizzate alcune funzionalità offerte da Sketch Engine, come per esempio Word Sketch, concordanze nel contesto e liste di collocazioni, che verranno illustrate più dettagliatamente nei paragrafi seguenti. Per prima cosa occorre chiarire cosa si intende con il termine collocazione, che frequentemente verrà utilizzato nel seguente capitolo dell'elaborato. Il termine collocazione può infatti essere definito come "[a] co-occurrence relationship between two words. Words are said to *collocate* with one another if one is more likely to occur in the presence of the other than elsewhere" (McEnery e Hardy, 2012:240). Un altro strumento che è stato utilizzato per analizzare i risultati dello studio è la funzione Word

²⁰http://www.fnopi.it/archivio_news/attualita/2093/Scheda%20n%20%202%20L-analisi%20dei%20dati%20Istat.pdf.

²¹<https://www.sketchengine.eu/ittenten-italian-corpus/#toggle-id-3>

²²<https://www.sketchengine.eu/>

Sketch²³ di Sketch Engine, uno strumento che permette di visualizzare in una singola pagina il comportamento grammaticale e collocazionale di una parola. I risultati sono organizzati in categorie, chiamate relazioni grammaticali, determinate secondo la grammatica Word Sketch (WSK), ovvero un insieme di regole che definiscono quali relazioni grammaticali e parole verranno mostrate dallo strumento Word Sketch. Sono inoltre state utilizzate liste di collocazioni ordinate secondo il valore LogDice, definito come “a statistic measure for identifying collocations. It expresses the typicality of the co-occurrence of the node and the collocates. It is only based on the frequency of the node and the collocates and the frequency of the whole collocation”²⁴. Infine sono state impiegate le concordanze nel contesto, ovvero una lista contenente tutti gli esempi della parola in analisi inserita nella frase da cui è stata estratta, mostrando quindi le porzioni di frase che circondano il termine o i termini in analisi. Il formato usato più frequentemente è chiamato concordanza KWIC, dove la parola in analisi è evidenziata al centro dello schermo ed è visualizzata una quantità uguale di contesto alla sua destra e alla sua sinistra²⁵.

²³ <https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/>

²⁴ https://www.sketchengine.eu/my_keywords/logdice/

²⁵ <https://www.sketchengine.eu/user-guide/concordance-introduction/>

Capitolo IV: Analisi dello studio

4.1 Presentazione dello studio

Nella tabella in appendice vengono riportate le frasi utilizzate per lo studio e l'output prodotto dai sistemi di traduzione automatica DeepL e Google Translate. Tutte le traduzioni sono state effettuate nel periodo dal 13/02/2020 al 14/02/2020. Le frasi sono state inserite nella finestra di traduzione dei due sistemi sia tutte insieme (con un limite di 5000 caratteri spazi inclusi alla volta), sia in quartetti di professioni (per un totale di 8 frasi per volta), sia soltanto una coppia di frasi per ogni professione (maschile/femminile), al fine di assicurarsi che le traduzioni rimanessero coerenti in presenza di ulteriori frasi e contesto. È emerso che l'output generato non subiva alcuna variazione. Ho proceduto quindi a riportare le frasi inglesi e gli output dei due sistemi di TA nelle tabelle in appendice (n.). Le professioni seguono l'ordine della tabella 1 riportata nel capitolo 3, contenente le professioni selezionate ordinate da quella con la minore presenza femminile a quella con la maggiore presenza femminile nell'occupazione. Sono presenti un totale di 112 frasi, costituite da una coppia di frasi per ognuna delle 56 professioni di cui 28 a maggioranza maschile e 28 a maggioranza femminile.

4.2 Analisi dei dati

4.2.1 Antecedente

L'antecedente, ovvero il pronome *him/her* viene sempre individuato correttamente da entrambi i sistemi di traduzione automatica. La frase "*I've known him/her for a long time*" viene infatti tradotta come "Lo/La conosco da molto tempo" a seconda della presenza di un pronome maschile o femminile nella frase. Nel totale di 112 frasi (56 frasi per le professioni a maggioranza maschile e 56 per le professioni a maggioranza femminile) l'antecedente è quindi sempre tradotto in maniera corretta.

4.2.2 Professione

Per verificare la correttezza della traduzione da parte dei sistemi di TA non è stato considerato solo il termine riferito alla professione, ma anche la traduzione del pronome dimostrativo *this*. Alcune professioni, infatti, vengono tradotte con termini

epiceni, ovvero che presentano la stessa forma per il maschile e il femminile, come per esempio le parole che terminano in -ista (es. dentista). La caratterizzazione maschile/femminile per queste parole viene determinata dagli articoli o dai pronomi che precedono il termine stesso, per questo motivo sono stati considerati corretti solo i termini che presentavano un accordo tra l'antecedente della frase (*him/her*) e il pronome dimostrativo. Per esempio è stata considerata corretta la traduzione “questa dentista” nelle frasi con antecedenti femminili, mentre non è stata considerata corretta la forma “questo nutrizionista” nelle frasi con antecedente femminile. L'analisi infatti è stata condotta da una prospettiva di accordo di genere, quindi seppure molti termini siano corretti dal punto di vista grammaticale o traduttivo, non lo sono dal punto di vista di genere o dell'uso di un linguaggio non sessista. L'analisi è stata condotta utilizzando la divisione tra professioni a maggioranza femminile e maschile, essendo 56 il numero totale delle professioni esaminate, suddivise in due gruppi da 28 professioni ciascuno, il risultato massimo è quindi di 28 per ognuno di loro. Mantenendo i set separati si può inoltre valutare il miglioramento (o il peggioramento) dell'output in base a quanto la frase possa essere considerata “stereotipica”, ovvero valutare se le professioni a maggioranza maschile vengono tradotte più frequentemente al maschile e viceversa. I due “subset” vengono definiti subset maschile (SM) e subset femminile (SF) per semplicità, ma si riferiscono alle professioni a maggioranza femminile o maschile, non al genere del soggetto della frase.

4.2.3 Subset professioni maschili

Nel SM tutte le professioni delle frasi con antecedente maschile (*him*) sono state correttamente identificate e tradotte da entrambi i sistemi di traduzione automatica. Nel caso delle frasi con antecedente femminile (*her*) invece, il numero di professioni tradotte correttamente si riduce in maniera drastica, sono infatti 5 su 28 le professioni tradotte correttamente da DeepL e 8 su 28 da Google Translate, come mostrato nella Figura 1.

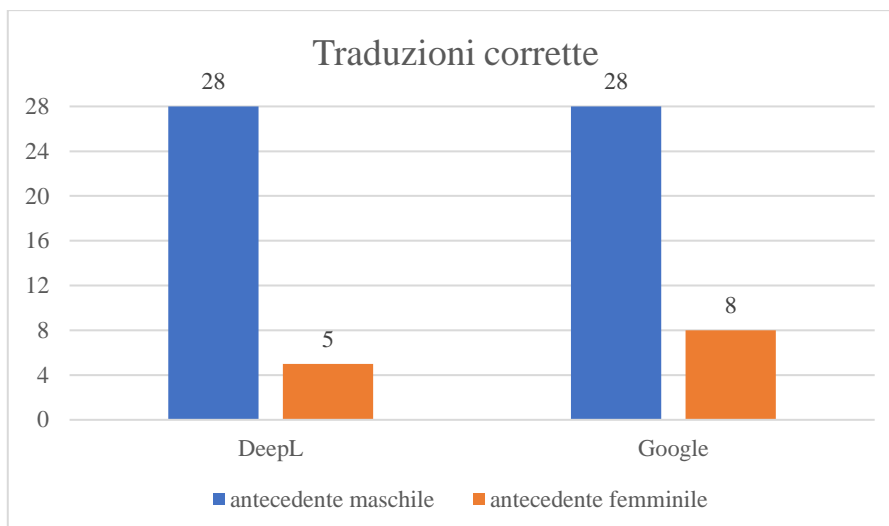


Figura 1: numero di professioni correttamente tradotte nel subset maschile

Tra queste occorre ricordare che è presente il termine *guida turistica*, che nella lingua italiana presenta solo la forma femminile, ed è stato quindi considerato corretto sia nelle frasi con antecedente maschile che femminile. Google Translate presenta inoltre un errore di traduzione, la frase “*I’ve known her for a long time, this butcher is one of the best in her field*” è stata tradotta come “La conosco da molto tempo, questa *macelleria* è una delle migliori nel suo campo”, in questo caso non è possibile valutare la correttezza della professione in sé in quanto il termine tradotto da GT non è la professione ma il negozio nel quale si svolge la stessa, per questo motivo è stato considerato errato. È stato inoltre considerata errata la traduzione della frase “*I’ve known her for a long time, this mail carrier is one of the best in her field*” con “La conosco da molto tempo, questo corriere è uno dei migliori nel suo campo”. Il motivo è che seppure non ci siano (al momento) attestazioni della forma *corriera* riferita alla professione e non al mezzo di trasporto e questa forma venga registrata dal dizionario Zingarelli²⁶ come rara, la lingua italiana prevede la forma femminile con suffisso in -*iera* per i nomi maschili che terminano in -*iere*, quindi sarebbe auspicabile utilizzare la forma *corriera*, nonostante possa essere considerata ambigua e per questo non sia largamente diffusa. Per lo stesso motivo non è stata considerata corretta la forma *medico* nelle frasi con antecedente femminile, perché; sebbene sia formalmente corretta e venga frequentemente utilizzata anche per riferirsi a donne (spesso con l’aggiunta della parola donna), sarebbe auspicabile utilizzare la forma *medica* per riferirsi alle donne che esercitano questa professione, come evidenziato sia dalle

²⁶ ristampa 2020 della dodicesima edizione

Raccomandazioni per un uso non sessista della lingua italiana di Sabatini (1987) sia da Robustelli (2017: online) in una consulenza linguistica per l'Accademia della Crusca.

Inoltre si può notare come i termini *police officer* e *firefighter*, siano tradotti con i termini poliziotto e pompiere, che, pur risultando corretti da un punto di vista formale, non lo sono dal punto di vista del registro né del punto di vista dell'uso più inclusivo della lingua. Come esposto nel capitolo 2 di questo elaborato infatti, nella lingua inglese sono stati abbandonati i termini *policeman* e *fireman* in quanto non inclusivi, sostituiti con i *police officer* e *firefighter*. In italiano i termini corrispondenti sono *agente di polizia* e *vigile del fuoco*, termini epiceni che risultano più inclusivi rispetto ai precedenti. Dal punto di vista del registro, i termini poliziotto e pompiere hanno un registro più familiare e colloquiale rispetto ad agente e vigile, che sono le denominazioni ufficiali delle professioni. Dallo studio emerge che le uniche professioni tradotte correttamente al femminile da entrambi i sistemi sono cantante (*singer*), scienziata (*scientist*), fotografa (*photographer*) e cuoca (*cook*), termine sul quale mi soffermerò nella sezione §4.3.5; inoltre GT traduce correttamente anche i termini pittrice (*painter*), dentista (*dentist*) e professoressa (*professor*). Facendo un confronto con i dati presentati nel capitolo III, si può notare che una sola professione tradotta correttamente al femminile appartiene al primo intervallo (0-20% di donne nella professione), 2 professioni appartengono al secondo intervallo (20-40% di donne nella professione) e 4 appartengono al terzo intervallo (40-60% di donne nella professione), possiamo quindi ipotizzare, pur sulla base di una quantità limitata di dati, che i termini al femminile siano più frequenti man mano che aumenta la percentuale di donne nella professione.

4.2.4 Subset professioni femminili

Per quanto riguarda SF, non tutte le professioni sono state individuate e correttamente tradotte. Nelle frasi con antecedente maschile (*him*) la professione viene tradotta correttamente in 24 casi su 28 da DeepL e in 26 su 28 da Google Translate, mentre nelle frasi con antecedente femminile (*her*) la professione viene tradotta in maniera corretta 17 volte da DeepL e 20 da GT.

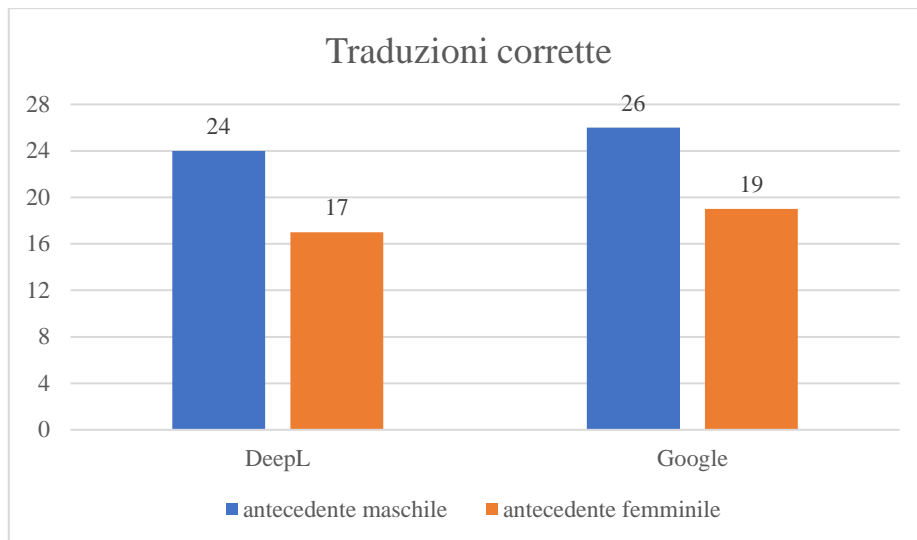


Figura 2: numero di professioni correttamente tradotte nel subset femminile

Google Translate presenta inoltre un errore di traduzione; infatti la frase “*I’ve known him for a long time, this cleaner is one of the best in his field*” viene tradotta come “lo conosco da molto tempo, questo *detergente* è uno dei migliori nel suo campo”, in questo caso non è possibile valutare la correttezza della professione in sé in quanto il termine tradotto da GT non si riferisce alla professione ma ad uno strumento utilizzato per pulire. È presente quindi un errore di traduzione simile a quello riscontrato nel subset maschile con il sostantivo *butcher*, erroneamente tradotto da GT con macelleria. È inoltre interessante analizzare la frase “*I’ve known him/her for a long time, this cleaner is one of the best in his/her field*”, che viene tradotta da DeepL come “Lo/la conosco da molto tempo, questa *donna delle pulizie* è una delle migliori nel suo campo” mentre Google fornisce una traduzione errata per la forma maschile, fornendo invece lo stesso output di DeepL per la forma femminile. La traduzione prodotta dai due sistemi è di fatto una forma non inclusiva, che si rivolge solo alle donne, ben lontana dal termine corretto, ovvero addetto/a alle pulizie. Inoltre in questo modo entrambi i sistemi contribuiscono ad alimentare lo stereotipo secondo cui tale professione sia esclusivamente femminile. Dagli output emerge inoltre come alcune professioni vengano tradotte sempre al femminile dai due sistemi, come nel caso di *nurse*, tradotto in tutti i casi come infermiera, declinando quindi al femminile la professione anche nelle frasi con antecedente maschile. Entrambe le professioni, seppure con molte differenze, rientrano nell’immaginario di professioni “femminili” proprio perché dedicate alla cura e all’igiene, tratti che sono stereotipicamente associati con il ruolo della donna. Si può infine notare come in questo set ci sia un

miglioramento generale delle traduzioni nelle frasi con antecedente femminile, e si può ipotizzare che sia dovuto anche alla maggiore presenza di termini al femminile per le professioni analizzate, nonché alla maggiore presenza di donne in queste professioni. Le professioni che non sono state declinate al femminile da nessuno dei due sistemi di TA in questo set sono: giudice (*judge*), magistrato (*magistrate*), fornaio/panettiere (*baker*), operatore sanitario (*health worker*) e sarto (*tailor*); inoltre DeepL non declina al femminile nutrizionista (*nutritionist*), artista (*artist*) e barista (*bartender*). Confrontando gli output dei due sistemi di TA con i dati sulla percentuale di donne nelle occupazioni selezionate (tabella n.), presentati nel capitolo 3, si può notare che una sola professione che non è stata tradotta al femminile da DeepL appartiene al quinto intervallo (80-100%), 2 professioni (non tradotte correttamente da entrambi i sistemi) appartengono al quarto intervallo (60-80%) e 5 appartengono al terzo intervallo (40-60%), possiamo quindi ipotizzare che i termini resi correttamente al femminile siano più frequenti man mano che aumenta la percentuale di donne nella professione.

4.2.5 Analisi sintagma nominale

Nella frase oggetto di studio è inoltre presente un sintagma nominale, ovvero “uno dei migliori nel suo campo” o “una delle migliori nel suo campo”, traduzione della frase “*one of the best in his/her field*”. Nei due subset viene infatti sempre identificato il pronome (*him/her*) presente nella prima frase, mentre la professione e il pronome dimostrativo non vengono sempre correttamente declinati al maschile o al femminile. Il sintagma nominale viene infatti accordato con la professione, non con l’antecedente (*him/her*) contenuto nel primo sintagma della frase e in aggiunta non viene considerato il pronome *him/her* presente nella frase originale in lingua inglese. Dall’analisi degli output emerge infatti come l’antecedente sia sempre stato identificato correttamente da entrambi i sistemi, con un totale di 56/56 occorrenze corrette sia nel subset maschile che femminile. Le professioni correttamente identificate e tradotte invece variano nei due subset, ma il sintagma nominale concorda sempre con la professione, sia nei casi in cui viene declinata correttamente sia in quelli in cui non viene utilizzata la forma corretta, come si può notare nella tabella 9.

		DeepL		Google Translate	
		Professioni	Sintagma nominale	Professioni	Sintagma nominale
SM	m	28/28	28/28	28/28	28/28
	F	5/28	5/28	8/28	9/28
SF	m	24/28	24/28	26/28	27/28
	F	17/28	17/28	19/28	19/28

Tabella 9: professioni e sintagmi nominali corretti

Sono stati considerati corretti i sintagmi che concordavano con il genere dell'antecedente *him/her*, ma si può notare come in due casi il numero di sintagmi nominali corretti di GT sia superiore al numero delle professioni correttamente declinate. La motivazione risiede nelle due traduzioni errate fornite nei subset da GT, il sistema infatti traduce *butcher* nel contesto femminile come macelleria e *cleaner* nel contesto maschile come detergente (come spiegato rispettivamente nei paragrafi §4.2.3 e §4.2.4 di questo capitolo) fornendo di fatto una traduzione errata non dal punto di vista del genere ma del significato. Considerando però l'accordo tra l'antecedente e il sintagma nominale nelle frasi in questione, il sintagma nominale appare corretto dal punto di vista del genere, e per questo motivo è stato inserito nel conteggio dei sintagmi nominali tradotti correttamente.

In aggiunta, il sintagma nominale analizzato suscita alcune riflessioni di tipo semantico e di interpretazione. Nella forma maschile, uno dei migliori nel suo campo, si intende che il soggetto è il migliore fra gli uomini e le donne della professione dato l'uso non marcato del maschile. D'altra parte la forma femminile, "una delle migliori nel suo campo" può essere interpretata sia nel senso "generico" in cui il soggetto è migliore fra uomini e donne, sia la frase può anche essere considerata in senso esclusivo, sottintendendo quindi che il soggetto è il migliore solo fra le donne della sua professione e non fra gli uomini. Questa interpretazione è causata da quelle che Francesco Sabatini definisce come "le limitazioni semantiche del femminile, conseguenti al fatto che è sempre marcato, mentre il maschile ha doppia valenza" (1987: 25). Il femminile è quindi considerato esclusivo, mentre il maschile può essere considerato in alcuni casi "inclusivo", anche se, come esposto nel capitolo 2 di questo elaborato, sono numerosi gli studi che si pongono in contrasto con questa interpretazione della lingua, preferendo sempre l'esplicitazione dei due generi. Tuttavia, ai fini del presente studio, i sintagmi nominali declinati al femminile sono

stati considerati corretti e quindi interpretati con un significato inclusivo. Questa riflessione pone nuovi spunti per continuare ad indagare in rapporto tra politiche linguistiche sempre più inclusive e traduzione automatica.

4.2.6 Confronto tra i due set e numero totale di errori

Confrontando le traduzioni delle frasi presenti nei due set possiamo notare come ci sia un forte sbilanciamento in favore del maschile. Infatti nel subset maschile le professioni con antecedente maschile vengono sempre tradotte in maniera corretta. Il dato non cambia nel subset contenente le professioni a maggioranza femminile, infatti anche in questo caso vengono maggiormente tradotte in maniera corretta le professioni con antecedente maschile rispetto a quelle con antecedente femminile. Le professioni vengono declinate al femminile in un numero inferiore di casi, anche in ambiti dove i dati statistici evidenziano una maggiore presenza di donne nella professione. Confrontando inoltre il totale delle frasi con antecedente maschile e quelle con antecedente femminile presenti in entrambi i subset, possiamo notare come la percentuale di errori vari notevolmente fra le frasi. Al fine di calcolare il numero totale di errori è stato considerato il totale delle frasi, nelle quali sono stati calcolati sia gli errori di traduzione della professione in base all'antecedente, sia gli errori relativi al mancato accordo tra il sintagma nominale e l'antecedente della frase in questione. Nella figura 3 vengono mostrati i risultati suddivisi in base al sistema e in base al genere dell'antecedente.

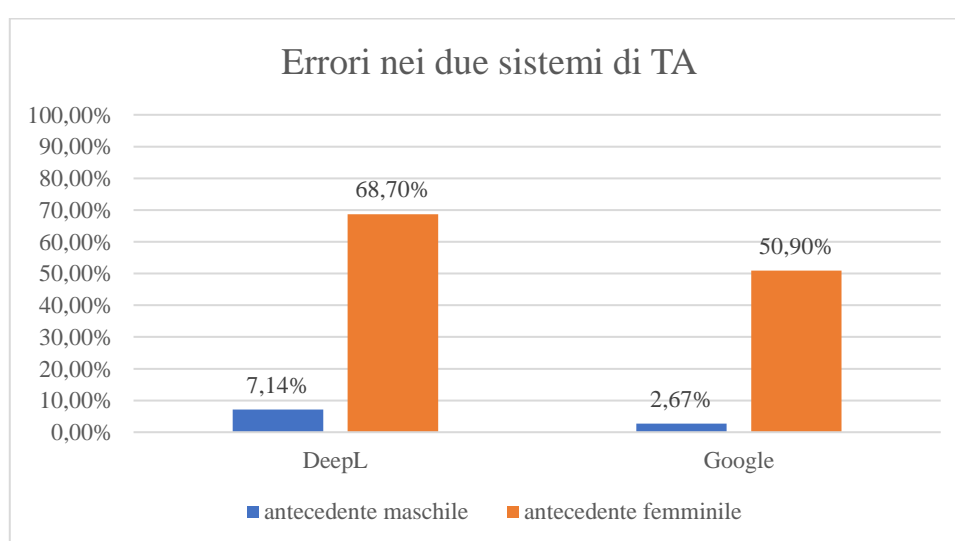


Figura 3: percentuale di errori delle professioni e del sintagma nominale nel due sistemi di TA

Appare evidente come le frasi con un antecedente femminile abbiano una percentuale di errori notevolmente superiore rispetto a quelle con antecedente maschile. Possiamo quindi affermare che i due sistemi, seppure in percentuali diverse, tendono a non identificare correttamente il femminile, generando quindi un aumento degli errori nell'output. Questo implica un maggiore sforzo in termini di post-editing, ovvero una quantità maggiore di tempo impiegato nell'attività di post-editing e di conseguenza maggiori costi dell'operazione. Questo potrebbe causare un aumento nell'uso del maschile generico per evitare errori di traduzione quando si usa la TA, compiendo quindi numerosi passi indietro rispetto all'uso più inclusivo della lingua e della maggiore visibilità del genere femminile.

4.3 Approfondimento sulle traduzioni dei termini chef/cuoco/cuoca

Un fenomeno interessante che emerge dallo studio e che merita un approfondimento è rappresentato dagli output generati dalle coppie di frasi *"I've known him/her for a long time, this chef is one of the best in his/her field"* e *"I've known him/her for a long time, this cook is one of the best in his/her field"*. Nel primo caso la professione viene sempre tradotta al maschile (questo chef) sia da DeepL che da Google Translate, mentre nel secondo caso la professione viene correttamente declinata al maschile (questo cuoco) e al femminile (questa cuoca) da entrambi i sistemi di traduzione automatica. Dato che si tratta di due professioni molto simili e che i due termini appartengono allo stesso campo semantico (e vengono in alcuni casi anche utilizzati come sinonimi) ho ritenuto utile soffermarmi maggiormente nell'analisi di queste frasi. Ho ipotizzato infatti che il termine *chef* abbia assunto un maggiore "prestigio sociale" e sia solitamente associato a professionisti di grande rilievo, generalmente maschi; mentre *cuoco* sia visto in maniera "inferiore" ma sempre in un contesto professionale e che infine il termine *cuoca* sia riferito principalmente all'attività casalinga della cucina, stereotipicamente appannaggio femminile. Per verificare la mia ipotesi ho quindi condotto uno studio utilizzando il corpus itTenTen16, composto da 4,9 miliardi di parole estratte da testi presenti sul web nel 2016 e consultato tramite la piattaforma SketchEngine.

4.3.1 Analisi dei collocati di chef

In primo luogo ho analizzato il numero di occorrenze del termine chef riferito al maschile o al femminile, e dato che la parola è epicena, e può di conseguenza essere usata in riferimento sia a uomini che a donne, ho aggiunto nella query CQL²⁷ anche articoli determinativi o indeterminativi, un aggettivo possessivo, una preposizione articolata o un pronome dimostrativo²⁸ in modo da far emergere in maniera chiara il genere della professione. Le ricerche effettuate hanno mostrato 3726 occorrenze per il sostantivo riferito al maschile e 1705 occorrenze al femminile, con quindi una netta prevalenza nell'uso del sostantivo chef riferito a uomini nel corpus.

Successivamente ho effettuato una ricerca del sostantivo chef e ho quindi analizzato le collocazioni del termine in uno span -3, 3; e nella figura 4 sono stati riportati i primi 20 risultati.

²⁷ The Corpus Query Language (CQL) is a special code or query language used in Sketch Engine to search for complex grammatical or lexical patterns (fonte: <https://www.sketchengine.eu/?s=cql>).

²⁸ CQL utilizzata: [lemma="lo" | lemma="uno" | lemma="dello" | lemma="nostro" | lemma="Lo" | lemma="Uno" | lemma="questo" | lemma="Questo"] [lemma="chef"] per la forma maschile e [lemma="la" | lemma="una" | lemma="della" | lemma="nostra" | lemma="La" | lemma="Una" | lemma="questa" | lemma="Questa"] [lemma="chef"] per la forma femminile.

	Word	Cooccurrences ?	Candidates ? ↓	LogDice
1	stellati	2,058	5,550	9.03 ***
2	stellato	2,194	16,494	9.01 ***
3	dallo	6,620	548,745	8.33 ***
4	executive	963	7,250	7.92 ***
5	Executive	931	13,730	7.80 ***
6	ristorante	2,671	275,463	7.78 ***
7	Rubio	635	2,239	7.37 ***
8	Ristorante	896	70,063	7.25 ***
9	Cracco	591	4,318	7.24 ***
10	Kenwood	586	4,356	7.23 ***
11	preparati	947	92,362	7.17 ***
12	cucina	3,463	685,315	7.13 ***
13	Chef	650	28,683	7.13 ***
14	patron	603	19,530	7.11 ***
15	sous	520	1,558	7.09 ***
16	piatti	1,593	266,360	7.07 ***
17	dello	11,906	2,868,089	7.03 ***
18	Bottura	488	4,029	6.97 ***
19	Michelin	499	17,528	6.86 ***
20	rinomati	480	14,503	6.83 ***

Figura 4: Elenco dei primi 20 collocati di chef, ordinati secondo il valore di LogDice.

Emerge come tra le parole che più frequentemente si associano a questa professione siano presenti termini come *stellato*, che indica uno dei maggiori riconoscimenti culinari, ovvero la stella Michelin; *executive* e *sous*, termini che indicano i più alti gradi delle brigate di cucina e numerosi cognomi o pseudonimi di famosi chef uomini, come per esempio Rubio (Gabriele Rubini), Cracco (Carlo) e Bottura (Massimo).

Vengono inoltre mostrati nella figura 5 i risultati ottenuti utilizzando la visualizzazione grafica di WordSketch, dove sono riportati i collocati più frequenti suddivisi per categorie, e per ogni categoria viene mostrato un massimo di 10 parole. La grandezza del cerchio nel quale la singola parola è iscritta rappresenta il maggiore o il minor valore della collocazione, quindi i cerchi di dimensioni maggiori rappresentano i termini che hanno un valore di LogDice maggiore.

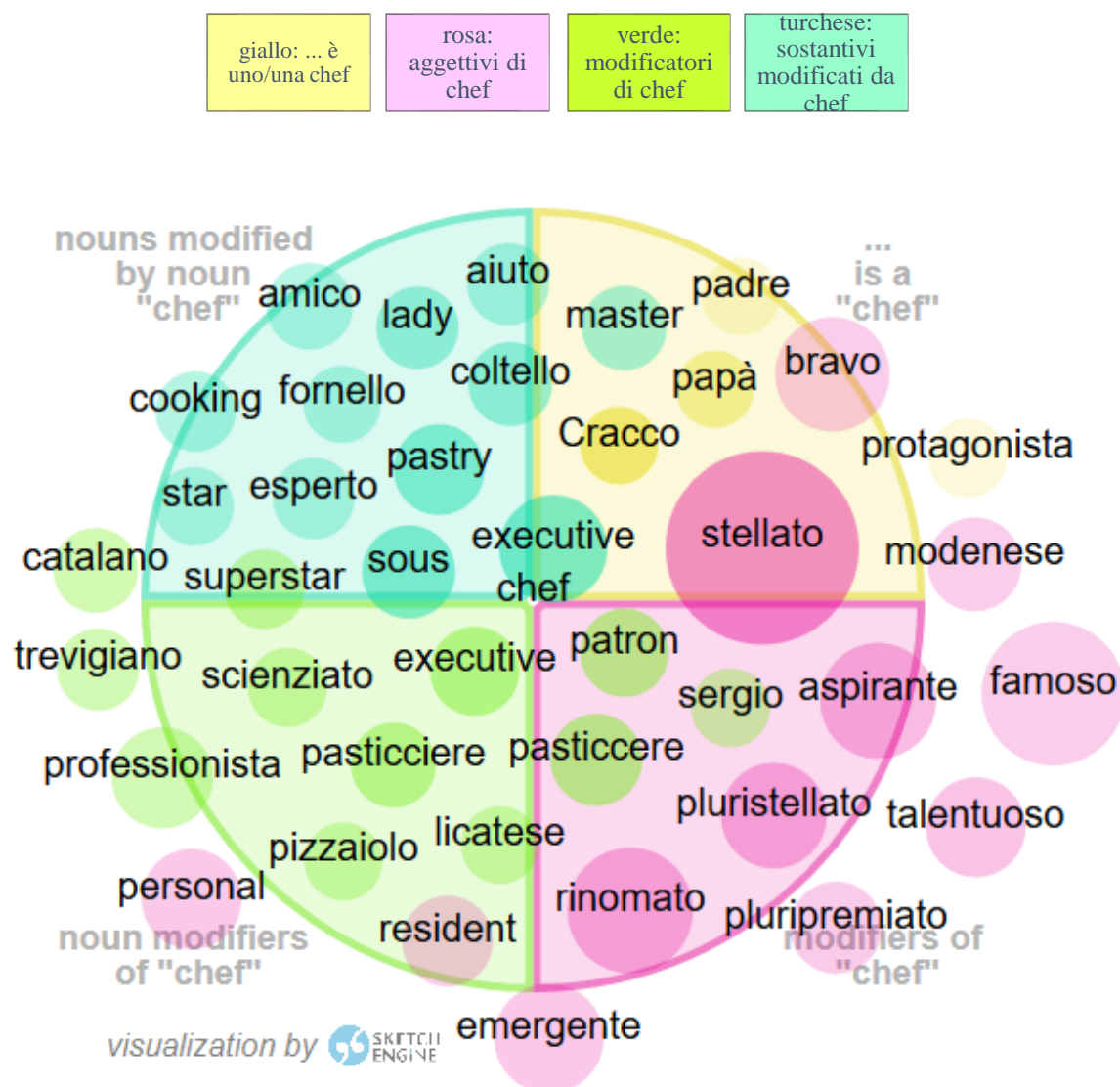


Figura 5: collocazioni del termine cuoco visualizzati tramite WordSketch

Come dall'analisi delle collocazioni elencate in figura 4, anche da questa visualizzazione si può notare come siano prevalenti i termini legati ai gradi più alti delle brigate di cucina, come *executive* e *sous*, ai riconoscimenti alla carriera, come *stellato*, *pluristellato* e *pluripremiato* o al successo e alla fama come *star*, *superstar* e *famoso*. Dall'analisi emerge inoltre la presenza, nella struttura grammaticale "...è uno/una chef" dei sostantivi *papà* e *padre*, che potrebbero essere ricollegati ad un ambito familiare. Per questo motivo ho analizzato nel contesto queste frasi, infatti nella struttura "... è un cuoco", i tre puntini rappresentano un sostantivo con funzione di soggetto della frase, e grazie alle funzioni di WordSketch è possibile visualizzare il contesto delle strutture rappresentate nella visualizzazione grafica semplicemente selezionando i risultati. Analizzando quindi le frasi nel contesto, emerge chiaramente

come il riferimento non sia ad un ambito familiare ma all'attività professionale, come si può vedere anche dalla figura 6.

vera, scattata per la realizzazione del libro "Mio papà è un vero	chef	"	</s><s> Foto di Giovanni Panarotto, scattata per il libro "Igles al
ati all'Università della Piza. </s><s> Figlio d'arte (il padre era lo	chef	del ristorante del Beatrice d'Este), propone pizze in due versioni.	
in questo contesto parlerà anche del suo libro "Il mio papà è uno	chef	". </s><s> Inoltre sempre i più piccoli, lunedì 25 ottobre , saranno	
so Channel (SKY). </s><s> Tre i libri all'attivo "Il mio papà è uno	chef	" (G. </s><s> Tommasi Editore, 2010); "Questo l'ho fatto io! </s><s>	
aso). </s><s> È genovese. </s><s> È figlio d'arte: il padre era lo	chef	di un noto ristorante del centro di Genova. </s><s> From Sardinia	
una famiglia dedita alla ristorazione, sia mamma che papà sono	chef	, e nella loro carriera hanno gestito qualcosa come 35 ristoranti. <	
onica che consiglio Giappone Il libro di cucina "Il mio papà è uno	chef	" di Francesca Romana Barberini </s><s> Camere vista mare per	
Saveve , del Ristorante da Antimo. </s><s> " Mio padre era uno	chef	e mia madre il suo aiuto cuoco, ma nella mia vita ho avuto la fortu	
I. </s><s> Sono entrambi figli d'arte: "Mio padre è un bravissimo	chef	, oggi docente all'Istituto alberghiero di Napoli", racconta Tony, e l	
e solo i piatti del proprio paese. </s><s> Anche mio padre è uno	chef	, ma non voleva che seguissi la sua strada, preferiva vedermi dot	
Romana Barberini. </s><s> Il suo primo libro "Il mio papà è uno	chef	" raccoglie le conversazioni con chef italiani sulla paternità e sull'e	
:/s><s> Un grande esempio per me. </s><s> Suo padre era uno	chef	come nel caso di Amelia Falco? </s><s> (Un'altra concorrente di	
nti e della ristorazione? </s><s> Figlio d'arte, suo padre è il noto	chef	Antonio Abbruzzino di Catanzaro, si potrebbe pensare che Luca :	

Figura 6: collocazioni e contesto del sostantivo chef

Ho inoltre verificato il numero di occorrenze della collocazione “chef stellato” e “chef stellata”; dato che sia dalla lista di collocazioni che dalla visualizzazione WordSketch emerge come stellato sia appunto il termine che più frequentemente viene associato alla parola in questione. Le occorrenze sono rispettivamente 3590 e 135, dimostrando ancora una volta una maggiore presenza del riferimento maschile nel corpus.

4.3.2 Analisi dei collocati di cuoco

In primo luogo sono state analizzate le collocazioni dei sostantivi cuoco sia tramite lista di collocazioni che tramite la visualizzazione grafica di WordSketch. Dalla lista delle prime 20 collocazioni più frequenti del sostantivo in analisi troviamo termini riferiti alla sfera professionale, come i sostantivi cameriere/i, pizzaiolo e pasticciere oppure legati ad un ambito di ricerca di lavoro, come cercasi.

	Word	Cooccurrences ?	Candidates ? ↓	LogDice
1	Prova	1,420	58,112	8.55 ***
2	camerieri	817	11,104	8.44 ***
3	cuoco	556	32,565	7.53 ***
4	Aiuto	456	17,716	7.48 ***
5	cameriere	358	19,517	7.10 ***
6	Federazione	899	148,783	7.10 ***
7	Cercasi	306	9,957	7.05 ***
8	aiuto	2,393	528,492	7.04 ***
9	chef	634	93,615	7.03 ***
10	AIUTO	276	3,593	7.03 ***
11	pizzaiolo	275	6,209	6.97 ***
12	pasticceri	253	4,710	6.88 ***
13	cuochi	274	21,088	6.69 ***
14	cuoche	210	2,608	6.65 ***
15	lavapiatti	202	2,520	6.60 ***
16	Italiana	1,065	308,823	6.54 ***
17	provetti	184	2,380	6.47 ***
18	stellati	181	5,550	6.38 ***
19	ristoratori	203	14,226	6.38 ***
20	Cuoco	181	6,047	6.37 ***

Figura 7: Lista delle prime 20 collocazioni del sostantivo cuoco

Nella Figura 8 sono invece riportati i collocati più frequenti suddivisi per categorie, e per ogni categoria viene mostrato un massimo di dieci parole e sono stati utilizzati gli stessi parametri impiegati anche nell'analisi del sostantivo chef.

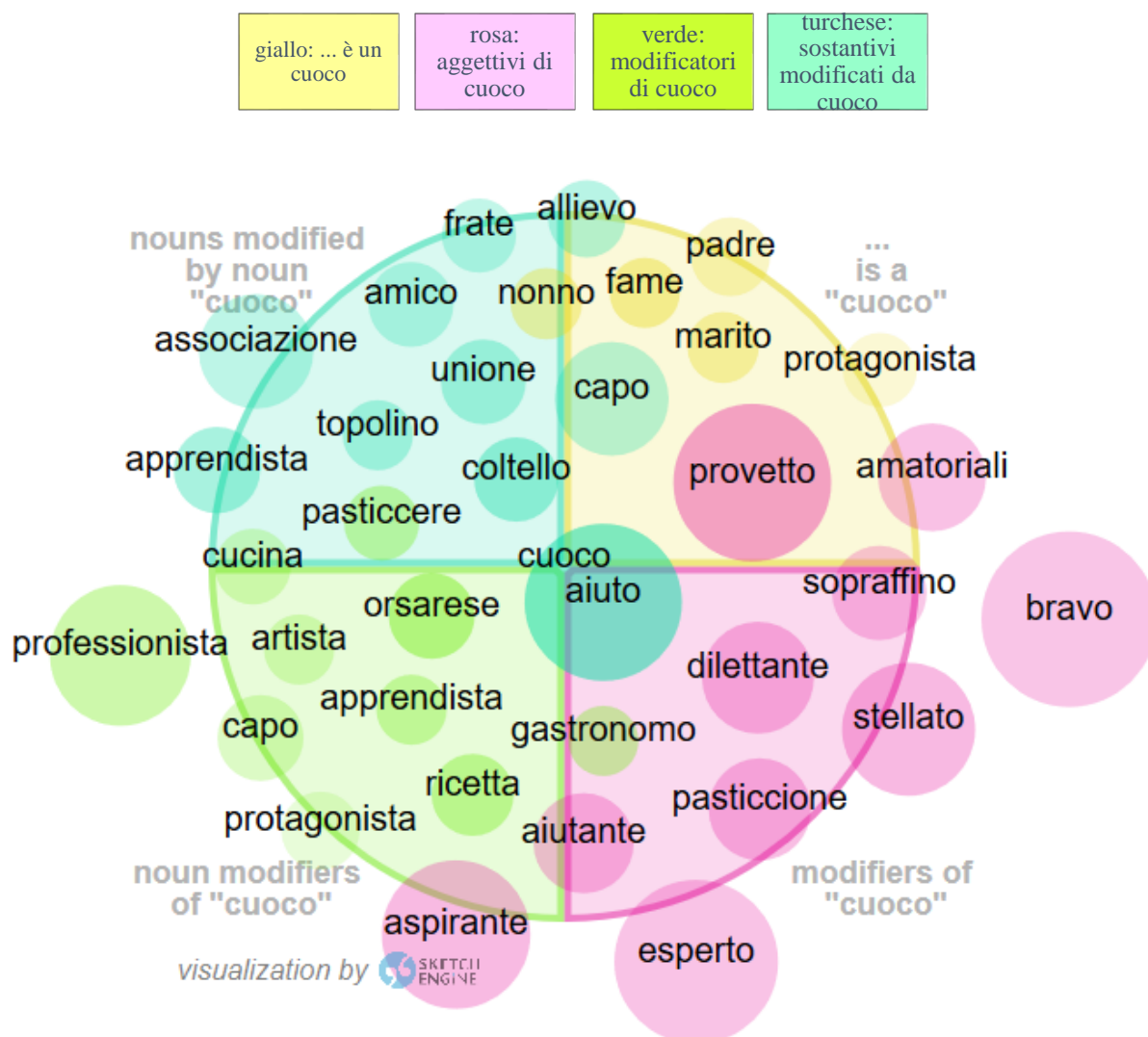


Figura 8:collocazioni del termine cuoco visualizzati tramite WordSketch

Nelle collocazioni di cuoco troviamo i termini pasticcere, professionista, ed esperto, elementi che rimandano ad un ambito professionale. Come evidenziato anche nell'analisi delle collocazioni del sostantivo chef presentata nella sezione §4.3.1, nella struttura "... è un cuoco", (dove i tre puntini rappresentano un sostantivo con funzione di soggetto della frase) troviamo i sostantivi nonno, padre e marito; che rimandano ad un ambito familiare e domestico. Al fine di analizzare nel dettaglio l'uso di questi sostantivi in questa specifica costruzione, come già esposto in precedenza per il sostantivo chef, ho analizzato le frasi nel contesto. Dall'approfondimento emerge come in realtà il riferimento sia sempre all'attività professionale della cucina, come dagli esempi riportati nella figura 9.

ta a terra con deviazione in angolo. </s><s> Mio **nonno** era un **cuoco** ovvero, l'esercizio della culinaria fra nostalgia e invenzione il m
di Cervia. </s><s> Delitto passionale? </s><s> La **vittima** è un **cuoco** 56enne. </s><s> L'arma del delitto potrebbe essere un'accetta. </s><s>
di stagione è il Bistrot dell'Alleanza, dove **protagonisti** sono i **cuochi** che in tutto il mondo si impegnano a proteggere la biodiversità </s><s>
artellotta: "Mi iscriverò all'alberghiero di Piacenza: mio **padre** è **cuoco** e voglio lavorare con lui". </s><s> Qualche spunto è arrivato ar
ontanara dei Lattari e del Faito " </s><s> I **protagonisti** sono i **cuochi** e le loro vite, le passioni, gli amori ... insomma i cuochi e quella
"ernando" ed è molto colorato. </s><s> I **protagonisti** sono un **cuoco** e una panettiera. </s><s> A leggerlo ci ho messo poco perché i
a verità è nel vino (in vino veritas). </s><s> La **fame** è il miglior **cuoco** che vi sia. </s><s> Mangiare è un bisogno dello stomaco, bere
le origini vicentine, Putin visse un'infanzia povera: il **nonno** era **cuoco** di Lenin e Stalin, il padre sommergibilista nella marina sovietica
[Articolo del 15-17 settembre 2014] </s><s> Mio **marito** è un **cuoco** Barilla. </s><s> Fa un sugo speciale. </s><s> Così ci ho chiest
blog sarà davvero un'esperienza interessante; mi **marito** è un **cuoco** e ho avuto a che fare con macchie davvero difficilissime. giusto

Figura 9: collocazioni e contesto del sostantivo cuoco

4.3.3 Analisi dei collocati di cuoca

Sono state poi analizzate le collocazioni del sostantivo cuoca, mostrate sia tramite una lista di collocazioni che utilizzando la visualizzazione grafica di WordSketch. Dalla lista delle prime 20 collocazioni più frequenti del sostantivo in analisi (mostrata in nella Figura 10) troviamo termini riferiti alla sfera domestica, come per esempio provetta, sopraffina e pasticciona.

	Word	Cooccurrences ?	Candidates ? ↓	LogDice
1	provetta	177	5,877	8.15 ***
2	cuochi	205	21,088	7.56 ***
3	sopraffina	85	2,566	7.35 ***
4	bravissima	186	28,204	7.15 ***
5	brava	364	70,019	7.14 ***
6	provette	66	2,985	6.95 ***
7	Pa	90	13,090	6.74 ***
8	cameriera	91	13,620	6.73 ***
9	esperte	92	21,348	6.39 ***
10	Ros	59	9,106	6.35 ***
11	pasticciona	38	1,097	6.31 ***
12	pasticcera	53	9,974	6.15 ***
13	cuoca	54	11,092	6.11 ***
14	bravissime	37	3,112	6.10 ***
15	cuoco	98	32,565	6.09 ***
16	esperta	118	42,790	6.08 ***
17	brave	50	13,739	5.86 ***
18	appassionata	74	30,216	5.76 ***
19	lavapiatti	25	2,520	5.58 ***
20	petulante	24	1,839	5.58 ***

Figura 10: Lista delle prime 20 collocazioni del sostantivo cuoca

Questi termini vengono solitamente utilizzati in situazioni informali o in un contesto casalingo. Dall'analisi possiamo inoltre notare che, a differenza degli altri due sostantivi in analisi, tra le collocazioni di cuoca troviamo un aggettivo con un significato negativo, ovvero petulante. Il termine in questione ha una connotazione negativa, e possiamo considerarlo un termine che stereotipicamente viene associato con maggiore frequenza alle donne rispetto agli uomini. È quindi interessante notare come già dai primi collocati sia presente un termine con valore spregiativo che invece non è presente nelle collocazioni del sostantivo cuoco o né in quelle di chef.

Nella Figura 11 sono invece riportati i collocati più frequenti suddivisi per categorie, e per ogni categoria viene mostrato un massimo di dieci parole utilizzando gli stessi parametri impiegati nell'analisi dei sostantivi chef e cuoco.

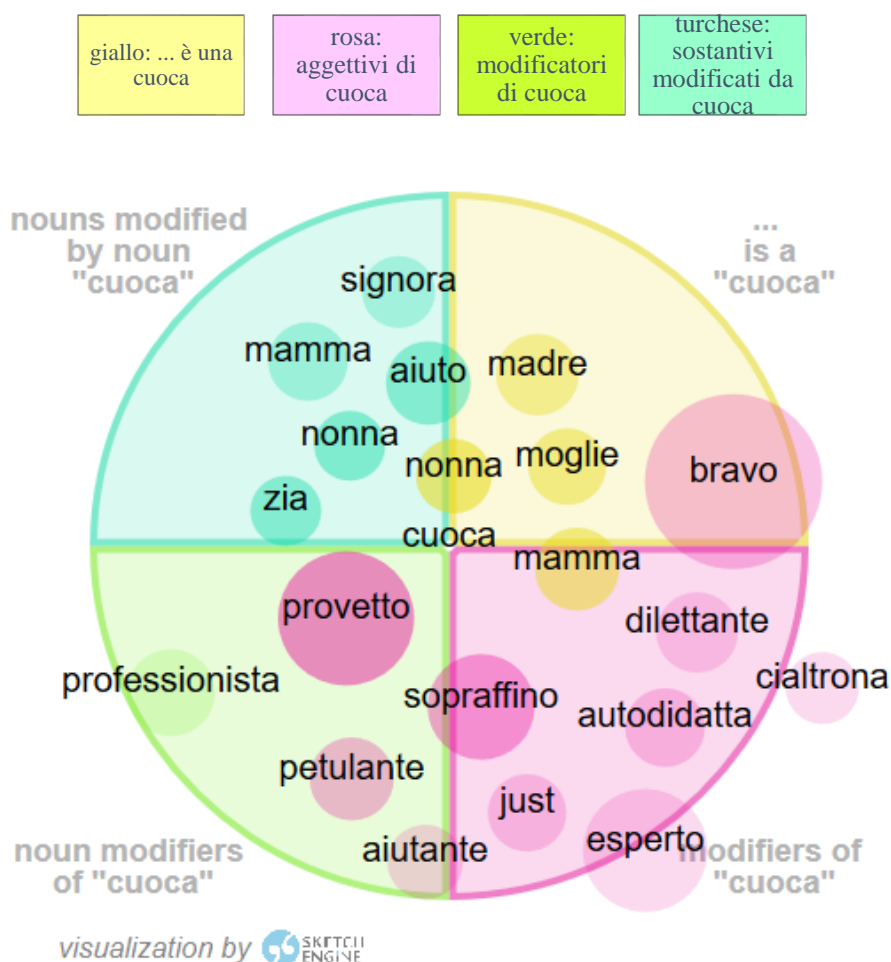


Figura 11: collocazioni del termine cuoca visualizzati tramite WordSketch

Tra le collocazioni di cuoca troviamo termini come cialtrona e petulante, che hanno una connotazione negativa, o i termini dilettante o provetta, che rimandano ad un ambito domestico e non professionale. Nella struttura "... è un cuoco" troviamo, come

nell'analisi dei sostantivi chef e cuoco presentato sopra, i sostantivi madre, moglie e nonna, che rimandano ad un ambito familiare e casalingo. Per verificare se anche in questo caso la struttura in realtà potesse riferirsi anche all'ambito professionale sono state analizzate nel dettaglio le frasi nel contesto, utilizzando lo stesso metodo impiegato nell'analisi dei due precedenti sostantivi, di cui si riportano alcuni esempi nella figura 12.

di quelli indimenticabili. </s><s> La nonna era una bravissima cuoca e per questa occasione speciale si era particolarmente impegna	
asa – anche nel cibo – siamo in par condicio: mia moglie è una cuoca formidabile dell'asado preparato in tanti modi, ma io vado ghiottc	
grande conoscitore di ricette. </s><s> Mia madre è la migliore cuoca in circolazione! </s><s> Quello che invece mi interessa molto è l	
io io sto cercando di spiegare a mio figlio che la mamma è una cuoca pessima e che alla mensa dell'asilo si mangia che è una favola!	
siamo realizzare il regalo su misura!! </s><s> La mamma è una cuoca provetta? una ballerina? una collezionista di bijoux? una donna :	
urarseler. </s><s> Frederik . Anche mia mamma era una brava cuoca , ma cucinava solo per la famiglia. </s><s> Per dessert ricordo u	
ione. </s><s> Tra le citazioni che preferisco: " Mia madre è una cuoca , e la fotografia per alcuni aspetti è come cucinare – un po' di ric	
IA, MANGIA TRANQUILLU !" </s><s> Quindi sua moglie è una cuoca priva di fantasia..... </s><s> Pirchi, a canusci? </s><s> A butt	
n soddisfava il tuo fine palato? </s><s> Lo so, la mamma è una cuoca eccezionale amore, non mangerai mai bene come a casa ma no	
nsomma ce ne sarà una che ti piace. </s><s> Mia madre è una cuoca esperta e io ho imparato prima ancora di camminare a far da ma	

Figura 12: collocazioni e contesto del sostantivo cuoca

In questo caso il riferimento è principalmente ad un ambito familiare, che rimanda alla cura della famiglia, attività che stereotipicamente viene attribuito alle donne. Non sono inoltre presenti riferimenti alla professione di cuoco, e di conseguenza l'atto di cucinare viene visto non come un'occupazione retribuita e professionale, ma come un'attività domestica.

Confrontando quindi le analisi dei collocati dei sostantivi chef, cuoco e cuoca si può dedurre che l'ipotesi iniziale appare corretta, di conseguenza i termini, nonostante possano essere utilizzati come sinonimi, non presentano significati e collocazioni sovrapponibili. Il termine chef viene infatti utilizzato in riferimento ad un ambito professionale di alto livello, assumendo quindi un maggior prestigio sociale. Il termine inoltre è utilizzato quasi esclusivamente al maschile nel corpus, uso che viene riscontrato anche negli output dei due sistemi di traduzioni automatica. Dall'analisi emerge inoltre come nella struttura "...è un" sono presenti per i sostantivi chef e cuoco termini come padre, papà o nonno e per il sostantivo cuoca i termini mamma, moglie o nonna, che potrebbero far pensare ad una somiglianza nel contesto d'uso di questi tre sostantivi. Tuttavia, come è stato esposto nelle sezioni precedenti, appare chiaro come nei primi due casi sia rivolto ad un contesto professionale, mentre nell'ultimo ad un contesto familiare e casalingo. Si può quindi ipotizzare che l'utilizzo del termine cuoca non mostri semplicemente un corretto utilizzo del femminile professionale, ma

come in realtà il termine sia presente al femminile perché stereotipicamente sono le donne a cucinare in casa, mentre gli uomini che cucinano svolgono questo compito in maniera professionale.

4.4 Approfondimento sulle traduzioni dei termini contabile/commercialista

Dall'analisi degli output emerge inoltre come la frase *“I’ve known him/her for a long time, this accountant is one of the best in his/her field”* venga resa utilizzando per le frasi con antecedente maschile come “Lo conosco da molto tempo, questo *commercialista* è uno dei migliori nel suo campo” (sia da DeepL che da GT), mentre viene resa con “La conosco da molto tempo, questa *contabile* è una dei migliori nel suo campo” nelle frasi con antecedente femminile (sia da DeepL che da Google Translate). I due termini non sono tuttavia sinonimi, infatti la professione di commercialista può essere esercitata solo da chi ha conseguito una laurea, mentre contabile, spesso utilizzato come sinonimo di ragioniere/a, si riferisce solitamente a un dipendente in possesso del solo diploma. Anche in questo caso, come già accaduto nel caso di chef/cuoco/cuoca esposto nel paragrafo precedente, la professione declinata al maschile ha un maggiore “prestigio sociale” rispetto a quella utilizzata per le frasi con antecedente femminile.

4.5 Approfondimento sull’output fornito da Google Translate

Vorrei inoltre soffermarmi in particolare sull’output proposto da Google Translate, considerando l’interesse che l’azienda ha dimostrato per la questione degli stereotipi di genere nel suo sistema di traduzione automatica. In un articolo pubblicato sul suo blog nel dicembre 2018²⁹, Google ha dichiarato di aver iniziato ad occuparsi della riduzione di stereotipi di genere nel suo sistema di traduzione automatica in molte combinazioni linguistiche, tra cui l’italiano. Se si utilizza il sistema come un dizionario, ovvero inserendo un solo termine nella finestra di traduzione, vengono fornite sia la traduzione al maschile che al femminile del termine in questione. Questa modalità, seppure non frequente tra linguisti e traduttori, potrebbe invece essere sfruttata maggiormente da un utente “non specializzato” o da apprendenti di lingua, come riportato anche in uno studio condotto da Eszter Tarsoly e Riitta-Liisa Valijärvi,

²⁹ <https://www.blog.google/products/translate/reducing-gender-bias-google-translate/>

dove veniva chiesto agli studenti di utilizzare Google Translate come un dizionario in esercizi grammaticali (2019: 94). Occorre segnalare però che questa modalità non è del tutto accurata: a fronte di numerosi termini corretti, come per esempio chirurga o rettrice, il sistema presenta presidentessa e avvocatessa come femminili delle parole presidente e avvocato, nonostante numerose linee guida propongano di evitare il suffisso -essa nei femminili.

Per questo motivo nella tabella 10 sono state riportate le 56 professioni selezionate per lo studio e l'output fornito da GT quando il termine viene inserito nella finestra di traduzione singolarmente, utilizzando il sistema proprio come se fosse un dizionario.

Termine inglese	Output
Electrician	Elettricista
Carpenter	Falegname
Firefighter	Pompiere
Pilot	Pilota
Painter (construction)	Pittrice/ pittore
Engineer (civil)	Ingegnere/ingegnera
Taxi driver	Tassista
Police officer	Poliziotto
Software developer	Sviluppatore di software
Programmer	Programmatrice/programmatore
Chef	Capocuoca/ capocuoco
Architect	Architetta/ architetto
Butcher	Macellaia/ macellaio
Chief executive	Amministratore delegato
Dentist	Dentista
Lawyer	Avvocatessa/avvocato
Mail carrier	Postino
Singer	Cantante
Janitor	Bidella/bidello
Physician	Medica/medico
Surgeon	Chirurga/chirurgo
Cook	Cuoco
Director	Direttrice/direttore

Tour guide	Guida turistica
Bus driver	Autista
Professor (post-secondary teacher)	Professoressa/ professore
Scientist (biological)	Scienziata/ scienziato
Photographer	Fotografa/ fotografo
Judge	Giudice
Magistrate	Magistrata/ magistrato
Bartender	Barista
Artist	Artist
Writer (technical)	Scrittrice/ scrittore
Baker	Panettiera/ panettiere
Pharmacist	Farmacista
Accountant	Contabile
Appraiser (claim)	Perita/ perito
Veterinarian	Veterinaria/ veterinario
Tax collector	Esattore delle tasse
Health worker	Operatore sanitario
Cashier	Cassiera/ cassiere
Model	Modella/ modello
Tailor	Sarta/ sarto
Psychologist	Psicologa/ psicologo
Librarian	Bibliotecaria/ bibliotecario
Teacher (elementary and middle school)	Insegnante
Flight attendant	Assistente di volo
Social worker	Assistente sociale
Teller	Cassiera/ cassiere
Occupational therapist	Terapista occupazionale
Nurse	Infermiera/ infermiere
Cleaner	Addetta alle pulizie/ addetto alle pulizie
Nutritionist	Dietologa/ dietologo
Hairdresser	Parrucchiera/ parrucchiere
Speech therapist	Logopedista
Dental hygienist	Igienista dentale

Lo scopo è quindi quello di mostrare come ci sia stato uno sforzo da parte dell'azienda nel mitigare o eliminare gli stereotipi di genere nel suo sistema di traduzione automatica, ma come in realtà queste modifiche non siano ancora effettive per tutti i sostantivi che denotano professioni. Un altro aspetto positivo che mostra una discreta sensibilità verso la questione di lingua e genere è la scelta, nei casi in cui vengono fornite le traduzioni del termine al maschile e al femminile, di anteporre quest'ultimo al maschile. In generale anche dai risultati emersi dallo studio Google Translate appare lievemente più accurato rispetto a DeepL nel declinare le professioni al maschile e al femminile, tuttavia appare evidente come alcuni termini correttamente tradotti quando viene utilizzato GT come dizionario non vengano poi riportati quando viene inserita una frase più complessa.

4.6 Discussione

Infine vorrei proporre una riflessione su alcuni termini che potremmo definire “di prestigio”, ovvero riferiti a professioni e/o ruoli apicali, categorie professionali nelle quali l'uso del femminile è ancora poco diffuso o accettato dai parlanti italiani. Nello studio sono infatti presenti i termini architetto/a, ingegnere/a, magistrato/, giudice, amministratore/amministratrice delegato/a, avvocato/a, chirurgo/a e medico/a, che rientrano nella definizione data e che sono stati tradotti nello studio solo nella forma maschile. Emerge quindi una forte disuguaglianza tra l'uso dei termini al maschile e al femminile di alcune professioni o ruoli apicali. Dato che il linguaggio ha un ruolo fondamentale nella nostra concezione della società, non utilizzare i femminili significa non dare visibilità alle donne che si trovano ai vertici della loro carriera. La mancanza di traduzioni corrette appare ancora più rilevante se si pensa che sempre più pagine web, tra cui notizie e post sui social media, vengono tradotte automaticamente senza alcun intervento di post-editing, contrastando di fatto tutti gli sforzi compiuti fino ad ora per adeguare la lingua italiana ad un uso non sessista. Infine lo sforzo compiuto per queste professioni andrebbe ampliato ad un numero sempre maggiore, normalizzando quindi l'uso al femminile di termini come perita, falegname o programmatrice, che sebbene abbiano (al momento) un numero di donne nella professione sicuramente inferiore rispetto ad altre, non devono essere per questo considerate meno importanti o degne di attenzioni. Dall'analisi dei dati emerge ancora

una forte disparità nell'uso di forme femminili per i sostantivi che denotano professioni o cariche da parte dei sistemi di traduzione automatica nonostante i livelli di accuratezza molto alti che è possibile ottenere dai sistemi neurali. Dai dati emerge inoltre come GT sia, seppur in maniera lieve, più accurato nelle traduzioni dei sostantivi relativi alla professione in frasi con antecedente femminile, dato che potrebbe essere ricondotto all'impegno dell'azienda nel mitigare la presenza di stereotipi di genere nel suo sistema di TA. Infine, è importante sottolineare che, nonostante gli sforzi compiuti da Google e l'impatto positivo che alcune implementazioni al suo sistema hanno comportato, il sistema presenta ancora numerosi stereotipi e non declina al femminile molte professioni, in particolare quelle relative ad ambiti stereotipicamente maschili o a incarichi e ruoli apicali. Questa analisi mostra quindi come sia necessario continuare ad indagare questi fenomeni e in particolare unire gli studi sull'elaborazione del linguaggio naturale ad una prospettiva di genere.

Conclusioni

Il presente elaborato si è concentrato su un ambito ancora poco studiato, che riguarda la presenza di stereotipi di genere nella traduzione automatica. Dopo aver fornito una panoramica generale sul funzionamento dei sistemi di traduzioni automatica e sulla questione del femminile delle professioni, che è uno degli ambiti dove è possibile notare con maggiore facilità la presenza di stereotipi, è stato elaborato e svolto uno studio di caso. L'argomento presentato in questo elaborato è molto recente, e di conseguenza anche la letteratura scientifica a disposizione è limitata ma in costante aumento, segno dell'interesse sempre crescente in materia. Nel capitolo I è stata definita la traduzione automatica, con particolare enfasi sulle architetture neurali, attualmente quelle più utilizzate. Sono poi stati esposti alcuni studi sugli stereotipi di genere nell'elaborazione del linguaggio naturale, soffermandosi su quelli nell'ambito della TA. Nel capitolo II è stato presentato brevemente il rapporto tra genere e lingua, in particolare il "caso" dei femminili professionali, sia in italiano che in altre lingue. Questa esposizione è servita per comprendere come lingue e paesi "affini" abbiano gestito la questione, in particolare illustrando i diversi approcci proposti dagli istituti nazionali che regolano o proteggono le rispettive lingue. Nel capitolo III sono stati presentati gli studi precedenti che hanno analizzato la presenza di stereotipi di genere nella traduzione automatica, al fine di elaborare una struttura frasale che permettesse di verificare la presenza di stereotipi tenendo però in considerazione le ambiguità della lingua italiana. È stata inoltre illustrata la metodologia scelta per la selezione delle professioni, in modo da rendere il campione statisticamente bilanciato. Sono inoltre stati presentati gli strumenti utilizzati nell'analisi *corpus-based* condotta nel capitolo successivo. Il Capitolo IV presenta infine un'accurata analisi degli output prodotti dai due sistemi di traduzione automatica utilizzati, DeepL e Google Translate, corredata da grafici illustrativi. In questo capitolo sono inoltre presenti degli approfondimenti su alcuni termini rilevanti e sulle traduzioni fornite da Google Translate.

Lo studio condotto in questa tesi presenta alcuni limiti che occorre menzionare. In primo luogo il campione di professioni e di frasi è molto limitato, e nonostante sia sufficiente per mostrare alcune caratteristiche della traduzione automatica in relazione con gli stereotipi di genere. In una prospettiva futura, infatti, potrebbe rivelarsi interessante aumentare il campione di professioni includendo anche quelle che non hanno dati disponibili o che sono particolarmente rilevanti per

un'analisi dal punto di vista della traduzione, come i ruoli politici e le cariche militari. Un altro spunto interessante, presentato anche in altri studi, potrebbe essere quello di indagare se e in quale misura alcuni aggettivi che stereotipicamente sono associati alle donne o agli uomini possano influenzare l'output traduttivo. La struttura frasale scelta permetterebbe di portare avanti questa ipotesi; inoltre la lista di aggettivi utilizzabili verrebbe creata tramite una ricerca basata su corpora, seguendo per esempio il modello di un precedente studio di Michael Pearce (2008) sul comportamento collocazionale delle parole "man" e "woman". Infine, il limite maggiore dello studio presentato è che le frasi sono state costruite *ad hoc* e non sono quindi attestate in nessun corpus. Dato che le frasi permettono di mettere in luce vari fenomeni, come per esempio l'accordo tra il coreferente e la professione o l'accordo tra aggettivo e professione, si potrebbe prendere in esame un solo fenomeno dei vari presenti nelle frasi e analizzarlo utilizzando frasi attestate. Ho ipotizzato, per esempio, che si potrebbe creare un corpus formato da frasi che contengano sia una professione che un riferimento (nome proprio o pronome) in grado di determinare il genere della professione. In questo modo si potrebbe creare un set di frasi attestate che possono essere tradotte utilizzando Google Translate e DeepL e confrontare gli output con quelli ottenuti dal set creato *ad hoc*.

Occorre precisare che queste sono solo alcune delle prospettive che gli studi in questo campo potrebbero intraprendere, e che la ricerca ha quindi un grandissimo potenziale di crescita. È rilevante però sottolineare l'importanza di unire l'ambito linguistico e traduttivo, le competenze in ambito computazionale e la prospettiva di genere al fine di offrire una panoramica più precisa e completa della questione. Molti degli studi sulla presenza di stereotipi di genere nell'elaborazione del linguaggio naturale, e in particolar modo nella traduzione automatica, mancano spesso di un'analisi del fenomeno traduttivo e linguistico in sé. Se da un lato gli studi con un maggior focus sulla componente informatica e computazionale della questione forniscono una visione dettagliata della traduzione automatica e dei sistemi che potrebbero mitigare o eliminare gli stereotipi di genere, dall'altro mancano spesso di considerazioni di carattere linguistico o sociologico/culturale. Seppur in un contesto "informatico" non dobbiamo dimenticare infatti che una lingua è indissolubilmente legata ad una cultura, e non può essere considerata come una mera variabile nella quale testare l'efficienza dell'intelligenza artificiale.

Bibliografia

Baeza-Yates, R. (2018). *Bias on the web*. Communications of the ACM, 61(6), pp.54-61.

Bahdanau, D., Van Merriënboer, B., Cho, K., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Disponibile da <https://www.aclweb.org/anthology/W14-4012/>

Becquer, A., & Institut national de la langue française (Francia). (1999). *Femme, j'écris ton nom: guide d'aide à la féminisation des noms des métiers, titres, grades et fonctions*. Paris: La Documentation Française.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3, 1137–1155.

Bhaskaran, J., & Bhallamudi, I. (2019). Good Secretaries, Bad Truck Drivers? Occupational Gender Stereotypes in Sentiment Analysis. *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, (pp. 66-68). Disponibile da <https://www.aclweb.org/anthology/W19-3809/>

Bollettieri Bosinelli R.M. (2005). Le prospettive di genere: discipline, soglie e confini. In Baccolini R. (2015). Bologna: Bononia University Press, pp 47-60.

Bosque, I. (2012). *Sexismo lingüístico y visibilidad de la mujer*. Madrid: Real Academia Española.

Chang, K. W., Prabhakaran, V., & Ordonez, V. (2019). Bias and Fairness in Natural Language Processing. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): Tutorial Abstracts*, Hong Kong, Cina. Disponibile da <https://www.aclweb.org/anthology/D19-2004/>

Costa-jussà, M. R. (2019). An analysis of gender bias studies in natural language processing. *Nature Machine Intelligence*, 1(11), 495–496. Disponibile da <https://www.nature.com/articles/s42256-019-0105-5>

Criado-Perez, C. (2019). *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Chatto & Windus.

Escudé Font, J. e M. R. Costa-Jussà (2019). Equalizing Gender Bias in Neural Machine Translation with Word Embeddings Technique, in Marta R. Costa-jussà, Christian Hardmeier, Will Radford, Kellie Webster (eds). *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, Florence, Italy: Association for Computational Linguistics, 147-15. <https://arxiv.org/pdf/1901.03116.pdf>.

Fioritto, A. (1993). *Codice di stile delle comunicazioni scritte ad uso delle amministrazioni pubbliche*. Roma: Istituto poligrafico e Zecca dello stato.

Fioritto, A. (1997). *Manuale di stile dei documenti amministrativi*. Bologna: Il Mulino.

Firth, J.R. (1968), *Selected papers of J. R. Firth, 1952-59*, a cura di Frank Robert Palmer London:Longmans,

Gaspari, F (2011). *Usare la traduzione automatica*. in Berselli, G. B.(eds). Bologna: CLUEB.

Ghenò, V. (2019). *Femminili singolari*. Firenze: Effequ.

Goldberg, Y., & Hirst, G. (2017). *Neural Network Methods in Natural Language Processing*. San Rafael: Morgan & Claypool.

Hecht-Nielsen, R., & Caudill, M. (1989). Neural Network Primer: Part I. *AI Expert*, 4(2), 61–67.

Hutchins, W. J., & Somers, H. L. (1992). *An introduction to machine translation*. London: Academic Press.

Jakubíček, M., Kilgariff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In *7th International Corpus Linguistics Conference CL* (pp. 125-127).

Koehn, P. (2017). Neural Machine Translation. In P. Koehn (Ed.), *Statistical Machine Translation* (2^a ed., pp. 5–101). Disponibile da <https://arxiv.org/abs/1709.07809>

Leavy, S. (2018). Gender bias in artificial intelligence. *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering - GE '18*, 14–16. Disponibile da <https://dl.acm.org/doi/proceedings/10.1145/3195570>

Levesque, H., Davis, E., L. Morgenstern (2012). The Winograd Schema Challenge. *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 553–561. Disponibile da <https://www.aaai.org/ocs/index.php/KR/KR12/paper/view/4492/0>

Lippmann, W. (1922). *Public Opinion*. New York: Harcourt, Brace & Company.

Lledó Cunill, E. (2006). *Las profesiones de la A a la Z, en femenino y en masculino* (Ed. riv.). Madrid: Instituto de la Mujer.

Luraghi, S., & Olita, A. (2006). *Linguaggio e genere: grammatica e usi*. Roma: Carocci.

Mazzara, B. M. (1997). *Stereotipi e pregiudizi*. Bologna: Il Mulino.

McEnery, T. & Hardie, A. (2012). *Corpus linguistics. Method, theory and practice*. Cambridge: Cambridge University Press.

Miller, C., & Swift, K. (1989). *The handbook of nonsexist writing: for writers, editors and speakers*. Londra: Women's Press.

Monti, J. (2017). Questioni di genere in traduzione automatica. In: *Al femminile, scritti linguistici in onore di Cristina Vallini*. De Meo, A., Di Pace, L., Manco, A., Monti J., & Pannain R. (Eds.). Firenze: Franco Cesati editore, 411-431.

Moulton, J. (1981) The Myth of the Neutral “Man”. In: Vetterling-Braggin, M., (Eds.), *Sexist Language*, Lanham: Rowman & Littlefield Pub Inc, 100-115

Mucchi Faina, A. (2006). *Comunicazione interculturale: il punto di vista psicologico-sociale*. Bari: GLF editori Laterza.

Palmonari, A., Cavazza, N., & Rubini, M. (2012). *Psicologia sociale* (2ª ed.). Bologna: Il Mulino.

Pearce M. (2008) Investigating the Collocational Behaviour of MAN and WOMAN in the BNC Using Sketch Engine, in *Corpora* 3 (1): 1–29. Disponibile da: <https://www.euppublishing.com/doi/full/10.3366/E174950320800004X>

Pettigrew, T. & Meertens, R. (1995). Subtle and Blatant Prejudice in Western Europe. *European Journal of Social Psychology*. 25. 57 - 75.

Prates, M., Avelar P. & Lamb L. (2019). Assessing Gender Bias in Machine Translation - A Case Study with Google Translate, in *Neural Computing and Application*: Springer London.

Robustelli, C. (2012) *Linee guida per l'uso del genere nel linguaggio amministrativo*. Firenze: Comune di Firenze.

Rudinger R., Naradowsky, J., Leonard B., & Van Durme B. (2018), Gender Bias in Coreference Resolution, In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, New Orleans, Louisiana: Association for Computational Linguistics, 8-14. Disponibile da <https://www.aclweb.org/anthology/N18-2002>

Sabatini A. (1987) *Il sessismo nella lingua italiana*, Roma: Istituto Poligrafico e Zecca dello Stato.

Sabatini F., (1987). *Più che una prefazione*, in Sabatini A. (1987), *Il sessismo nella lingua italiana*, Roma: Istituto Poligrafico e Zecca dello Stato, pp. 13-22.

Stanovsky, G., Smith, N. A. & Zettlemoyer L. (2019) Evaluating Gender Bias in Machine Translation, In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy.

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP*. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). Disponibile da <https://www.aclweb.org/anthology/events/acl-2019/>

Tajfel, H. (1981). *Gruppi umani e categorie sociali*. Bologna: Il Mulino (trad.it. 1995)

- Toral, A., Wieling, M., & Way, A. (2018). Post-editing Effort of a Novel with Statistical and Neural Machine Translation. *Frontiers in Digital Humanities*, 5. Disponibile da <https://www.frontiersin.org/articles/10.3389/fdigh.2018.00009/full>
- Valijärvi, R.L. e Tarsoly, E. (2019) 'Language students as critical users of Google Translate': pitfalls and possibilities, *Practitioner Research in Higher Education Journal*, 12(1), pp. 61-74. Disponibile da: <https://ojs.cumbria.ac.uk/index.php/prhe/article/view/511>
- Villano, P. (2003). *Pregiudizi e stereotipi*. Roma: Carocci.
- Weaver, W. (1949). "Memorandum". Disponibile online: <http://www.mt-archive.info/Weaver-1949.pdf>
- Zarra, G., & Gomez Gane, Y. (2017). *"Quasi una rivoluzione": i femminili di professioni e cariche in Italia e all'estero*. Firenze: Accademia della Crusca.
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. (2018). Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, New Orleans, Louisiana: Association for Computational Linguistics, 15–20. Disponibile da <https://www.aclweb.org/anthology/N18-2003/>
- Zingarelli, N. (2020). *Lo Zingarelli. Vocabolario della lingua italiana*. Bologna: Zanichelli.

Sitografia

Académie Française, (2012) “Féminisation des noms de métiers, fonctions, grades et titres”. <https://web.archive.org/web/20181122152133/http://www.academie-francaise.fr/actualites/feminisation-des-noms-de-metiers-fonctions-grades-et-titres>

[ultima consultazione: 12/01/2020]

Académie Française, (2014) “La féminisation des noms de métiers, fonctions, grades ou titres - Mise au point de l’Académie Française.” <http://www.academie-francaise.fr/actualites/la-feminisation-des-noms-de-metiers-fonctions-grades-ou-titres-mise-au-point-de-lacademie> [ultima consultazione: 13/01/2020]

Académie Française, (2019) “La féminisation des noms de métiers et de fonctions.” <http://www.academie-francaise.fr/actualites/la-feminisation-des-noms-de-metiers-et-de-fonctions> [ultima consultazione: 14/01/2020]

Bazzanella, C. (2010) “Genere e lingua”. [http://www.treccani.it/enciclopedia/genere-e-lingua_\(Enciclopedia-dell'Italiano\)/](http://www.treccani.it/enciclopedia/genere-e-lingua_(Enciclopedia-dell'Italiano)/) [7/01/2020]

Bearne, S. (2016) “From air hostesses to firemen ...it's time to drop gendered job titles”. The Guardian, 5 maggio 2016

<https://www.theguardian.com/careers/2016/may/05/from-air-hostesses-to-firemen-its-time-to-drop-gendered-job-titles> [ultima consultazione: 28/01/2020]

Censis, (2018). “Percorsi e scenari dell’avvocatura italiana: rapporto 2018”. www.cassaforense.it/media/7194/rapporto-censis-2018.pdf [ultima consultazione 18/02/2020]

Consiglio Nazionale Ordine Psicologi (2018). “Dati statistici”. <https://www.psy.it/dati-statistici> [ultima consultazione 18/02/2020]

Consiglio Superiore della Magistratura, (2019). “Donne in magistratura”. <http://bit.ly/2ZD6cSZ> [ultima consultazione 18/02/2020]

Diño, G. (2019) “He Said, She Said: Addressing Gender in Neural Machine Translation”. <https://slator.com/technology/he-said-she-said-addressing-gender-in-neural-machine-translation/> [ultima consultazione: 5/11/2019]

Dipartimento di Interpretazione e Traduzione, (2020). “Ricercatrici, assegniste, professoressa?”. <https://dit.unibo.it/it/dipartimento/persone> [ultima consultazione: 16/02/2020]

Federazione Nazionale degli Ordini dei Medici Chirurghi e degli Odontoiatri, (2019). “8 marzo: la medicina è donna, l’odontoiatria tiene il passo”. <https://portale.fnomceo.it/8-marzo-la-medicina-e-donna-lodontoiatria-tiene-il-passo/> [ultima consultazione: 18/02/2020]

Federazione Nazionale Ordini Professioni Infermieristiche, (2016). “Tutti i numeri degli infermieri. Chi sono, dove lavorano, privati, dipendenti e disoccupati: una professione allo specchio” [.http://www.fnopi.it/archivio_news/attualita/2093/Scheda%20n%202%20L-analisi%20dei%20dati%20Istat.pdf](http://www.fnopi.it/archivio_news/attualita/2093/Scheda%20n%202%20L-analisi%20dei%20dati%20Istat.pdf). [ultima consultazione 18/02/2020]

Federazione nazionale Ordini Veterinari Italiani, (2019). “33.300 medici veterinari in Italia, in crescita le donne ma non i guadagni”. <https://www.fnovi.it/node/47990> [ultima consultazione 18/02/2020]

Federazione Nazionale unitaria titolari di farmacia, (2019). “Numero di farmacisti”. <https://www.federfarma.it/Farmaci-e-farmacie/Notizie-e-dati-dall-Europa/Numero-di-farmacisti.aspx> [ultima consultazione 18/02/2020]

Gehring, J. (2017). “A novel approach to neural machine translation” <https://engineering.fb.com/ml-applications/a-novel-approach-to-neural-machine-translation/> [ultima consultazione: 28/10/2019]

González Ferrán, J, (2017). “«Profesora», «médica», «jueza», «pilota»..., profesiones en femenino”. <https://www.fundeu.es/noticia/profesora-medica-jueza-pilota-profesiones-en-femenino/> [ultima consultazione: 15/01/2020]

Gruppo di lavoro per un approccio di genere al linguaggio amministrativo e istituzionale. (2017). “Generi e linguaggi. Linee guida per un linguaggio amministrativo e istituzionale attento alle differenze di genere.” <https://www.unipd.it/node/48451> [ultima consultazione: 18/12/2019]

Haddow, B. (2016) “Health In My Language. Making public health information available in a wider variety of languages”. <http://www.himl.eu/> [ultima consultazione 30/10/2019]

Kuczmariski, J. (2018) “Reducing gender bias in Google Translate”.
<https://www.blog.google/products/translate/reducing-gender-bias-google-translate/>
[ultima consultazione: 20/11/2019]

Ministerio de Educación y Ciencia (1988) “Recomendaciones para el uso no sexista de la lengua” <https://sede.educacion.gob.es/publiventa/recomendaciones-para-el-uso-no-sexista-de-la-lengua-coeducacion/igualdad/21177> [ultima consultazione 15/01/2020]

Ministero dell’Interno, (2019). “Presenza delle donne nel ministero dell’Interno”.
<https://www.interno.gov.it/it/sala-stampa/dati-e-statistiche/presenza-donne-nel-ministero-dellinterno> [ultima consultazione 18/02/2020]

Ministero dell’Istruzione, dell’Università e della Ricerca, (2019) “Docenti a tempo indeterminato. Scuola statale”
<https://dati.istruzione.it/opendata/opendata/catalogo/elements1/?area=Personale%20Scuola> [ultima consultazione 18/02/2020]

Ministero dell’Istruzione dell’Università e della Ricerca, (2019). “Focus sul personale docente e non docente nel sistema universitario italiano - a.a. 2017/2018”.
<http://ustat.miur.it/documenti/2019-notiziario-statistico-n-2/> [ultima consultazione 18/02/2020]

O’Brien, S. (2019). “The INTERnAtional network on Crisis Translation”.
<https://cordis.europa.eu/project/id/734211> [ultima consultazione 30/10/2019]

Robustelli, C. (2017). “Donne al lavoro (medico, direttore, poeta): ancora sul femminile dei nomi di professione”.
<https://accademiadellacrusca.it/it/consulenza/donne-al-lavoro-medico-direttore-poeta-ancora-sul-femminile-dei-nomi-di-professione/1237> [ultima consultazione: 25/11/2019]

Sketch Engine, (2019). “Concordance – a list of examples in context”.
<https://www.sketchengine.eu/user-guide/concordance-introduction/> [ultima consultazione 23/02/2020]

Sketch Engine, (2019). “CQL – Corpus Query Language”.
<https://www.sketchengine.eu/?s=cql> [ultima consultazione 23/02/2020]

Sketch Engine, (2019). “itTenTen: Corpus of the Italian Web”.
<https://www.sketchengine.eu/ittenten-italian-corpus/#toggle-id-3> [ultima consultazione 23/02/2020]

Sketch Engine, (2019). “Log Dice”.
https://www.sketchengine.eu/my_keywords/logdice/ [ultima consultazione 23/02/2020]

Sketch Engine, (2019). “Word Sketch — collocations and word combinations”.
<https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/>. [ultima consultazione 23/02/2020]

Stanovsky, G. (2019). “Evaluating Gender Bias in Machine Translation”
https://github.com/gabrielStanovsky/mt_gender [ultima consultazione 15/12/2019]

The Stanford Natural Language Processing Group, (2016). “Coreference Resolution”.
<https://nlp.stanford.edu/projects/coref.shtml> [ultima consultazione 15/12/2019]

US Bureau of Labor Statistics, (2020). “Labor Force Statistics from the Current Population Survey” <https://www.bls.gov/cps/cpsaat11.htm> [ultima consultazione 12/02/2020]

Uszkoreit, J (2017). “Transformer: A Novel Neural Network Architecture for Language Understanding” <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html> [ultima consultazione 28/10/2019]

Vilella, P. (2012) “¿Es sexista el idioma español?”, BBC Mundo, 8 marzo 2018.”
https://www.bbc.com/mundo/noticias/2012/03/120308_sexismo_idioma_espanol.shtml [ultima consultazione: 15/02/2020]

Allegati

Input originale (subset maschile)	Output DeepL	Output Google
I've known him for a long time, this electrician is one of the best in his field.	Lo conosco da molto tempo, questo elettricista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo elettricista è uno dei migliori nel suo campo.
I've known her for a long time, this electrician is one of the best in her field.	La conosco da molto tempo, questo elettricista è uno dei migliori nel suo campo.	La conosco da molto tempo, questo elettricista è uno dei migliori nel suo campo.
I've known him for a long time, this carpenter is one of the best in his field.	Lo conosco da molto tempo, questo falegname è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo falegname è uno dei migliori nel suo campo.
I've known her for a long time, this carpenter is one of the best in her field.	La conosco da molto tempo, questo falegname è uno dei migliori nel suo campo.	La conosco da molto tempo, questo carpentiere è uno dei migliori nel suo campo.
I've known him for a long time, this firefighter is one of the best in his field.	Lo conosco da molto tempo, questo pompiere è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo pompiere è uno dei migliori nel suo campo.
I've known her for a long time, this firefighter is one of the best in her field.	La conosco da molto tempo, questo pompiere è uno dei migliori nel suo campo.	La conosco da molto tempo, questo pompiere è uno dei migliori nel suo campo.
I've known him for a long time, this pilot is one of the best in his field.	Lo conosco da molto tempo, questo pilota è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo pilota è uno dei migliori nel suo campo.

I've known her for a long time, this pilot is one of the best in her field.	La conosco da molto tempo, questo pilota è uno dei migliori nel suo campo.	La conosco da molto tempo, questo pilota è uno dei migliori nel suo campo.
I've known him for a long time, this painter is one of the best in his field.	Lo conosco da molto tempo, questo pittore è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo pittore è uno dei migliori nel suo campo.
I've known her for a long time, this painter is one of the best in her field.	La conosco da molto tempo, questo pittore è uno dei migliori nel suo campo.	La conosco da molto tempo, questa pittrice è una delle migliori nel suo campo.
I've known him for a long time, this engineer is one of the best in his field.	Lo conosco da molto tempo, questo ingegnere è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo ingegnere è uno dei migliori nel suo campo.
I've known her for a long time, this engineer is one of the best in her field.	La conosco da molto tempo, questo ingegnere è uno dei migliori nel suo campo.	La conosco da molto tempo, questo ingegnere è uno dei migliori nel suo campo.
I've known him for a long time, this taxi driver is one of the best in his field.	Lo conosco da molto tempo, questo tassista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo tassista è uno dei migliori nel suo campo.
I've known her for a long time, this taxi driver is one of the best in her field.	La conosco da molto tempo, questo tassista è uno dei migliori nel suo campo.	La conosco da molto tempo, questo tassista è uno dei migliori nel suo campo.
I've known him for a long time, this police officer is one of the best in his field.	Lo conosco da molto tempo, questo poliziotto è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo poliziotto è uno dei migliori nel suo campo.

I've known her for a long time, this police officer is one of the best in her field.	La conosco da molto tempo, questo poliziotto è uno dei migliori nel suo campo.	La conosco da molto tempo, questo poliziotto è uno dei migliori nel suo campo.
I've known him for a long time, this software developer is one of the best in his field.	Lo conosco da molto tempo, questo sviluppatore di software è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo sviluppatore di software è uno dei migliori nel suo campo.
I've known her for a long time, this software developer is one of the best in her field.	La conosco da molto tempo, questo sviluppatore di software è uno dei migliori nel suo campo.	La conosco da molto tempo, questo sviluppatore di software è uno dei migliori nel suo campo.
I've known him for a long time, this programmer is one of the best in his field.	Lo conosco da molto tempo, questo programmatore è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo programmatore è uno dei migliori nel suo campo.
I've known her for a long time, this programmer is one of the best in her field.	La conosco da molto tempo, questo programmatore è uno dei migliori nel suo campo.	La conosco da molto tempo, questo programmatore è uno dei migliori nel suo campo.
I've known him for a long time, this chef is one of the best in his field.	Lo conosco da molto tempo, questo chef è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo chef è uno dei migliori nel suo campo.
I've known her for a long time, this chef is one of the best in her field.	La conosco da molto tempo, questo chef è uno dei migliori nel suo campo.	La conosco da molto tempo, questo chef è uno dei migliori nel suo campo.

I've known him for a long time, this architect is one of the best in his field.	Lo conosco da molto tempo, questo architetto è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo architetto è uno dei migliori nel suo campo.
I've known her for a long time, this architect is one of the best in her field.	La conosco da molto tempo, questo architetto è uno dei migliori nel suo campo.	La conosco da molto tempo, questo architetto è uno dei migliori nel suo campo.
I've known him for a long time, this butcher is one of the best in his field.	Lo conosco da molto tempo, questo macellaio è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo macellaio è uno dei migliori nel suo campo.
I've known her for a long time, this butcher is one of the best in her field.	La conosco da molto tempo, questo macellaio è uno dei migliori nel suo campo.	La conosco da molto tempo, questa macelleria è una delle migliori nel suo campo.
I've known him for a long time, this chief executive is one of the best in his field.	Lo conosco da molto tempo, questo amministratore delegato è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo amministratore delegato è uno dei migliori nel suo campo.
I've known her for a long time, this chief executive is one of the best in her field.	La conosco da molto tempo, questo amministratore delegato è uno dei migliori nel suo campo.	La conosco da molto tempo, questo amministratore delegato è uno dei migliori nel suo campo.
I've known him for a long time, this dentist is one of the best in his field.	Lo conosco da molto tempo, questo dentista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo dentista è uno dei migliori nel suo campo.

I've known her for a long time, this dentist is one of the best in her field.	La conosco da molto tempo, questo dentista è uno dei migliori nel suo campo.	La conosco da molto tempo, questa dentista è una delle migliori nel suo campo.
I've known him for a long time, this lawyer is one of the best in his field.	Lo conosco da molto tempo, questo avvocato è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo avvocato è uno dei migliori nel suo campo.
I've known her for a long time, this lawyer is one of the best in her field.	La conosco da molto tempo, questo avvocato è uno dei migliori nel suo campo.	La conosco da molto tempo, questo avvocato è uno dei migliori nel suo campo.
I've known him for a long time, this mail carrier is one of the best in his field.	Lo conosco da molto tempo, questo postino è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo corriere è uno dei migliori nel suo campo.
I've known her for a long time, this mail carrier is one of the best in her field.	La conosco da molto tempo, questo postino è uno dei migliori nel suo campo.	La conosco da molto tempo, questo corriere è uno dei migliori nel suo campo.
I've known him for a long time, this singer is one of the best in his field.	Lo conosco da molto tempo, questo cantante è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo cantante è uno dei migliori nel suo campo.
I've known her for a long time, this singer is one of the best in her field.	La conosco da molto tempo, questa cantante è una delle migliori nel suo campo.	La conosco da molto tempo, questa cantante è una delle migliori nel suo campo.

I've known him for a long time, this janitor is one of the best in his field.	Lo conosco da molto tempo, questo bidello è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo bidello è uno dei migliori nel suo campo.
I've known her for a long time, this janitor is one of the best in her field.	La conosco da molto tempo, questo bidello è uno dei migliori nel suo campo.	La conosco da molto tempo, questo bidello è uno dei migliori nel suo campo.
I've known him for a long time, this physician is one of the best in his field.	Lo conosco da molto tempo, questo medico è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo medico è uno dei migliori nel suo campo.
I've known her for a long time, this physician is one of the best in her field.	La conosco da molto tempo, questo medico è uno dei migliori nel suo campo.	La conosco da molto tempo, questo medico è uno dei migliori nel suo campo.
I've known him for a long time, this surgeon is one of the best in his field.	Lo conosco da molto tempo, questo chirurgo è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo chirurgo è uno dei migliori nel suo campo.
I've known her for a long time, this surgeon is one of the best in her field.	La conosco da molto tempo, questo chirurgo è uno dei migliori nel suo campo.	La conosco da molto tempo, questo chirurgo è uno dei migliori nel suo campo.
I've known him for a long time, this cook is one of the best in his field.	Lo conosco da molto tempo, questo cuoco è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo cuoco è uno dei migliori nel suo campo.

I've known her for a long time, this cook is one of the best in her field.	La conosco da molto tempo, questa cuoca è una delle migliori nel suo campo.	La conosco da molto tempo, questa cuoca è una delle migliori nel suo campo.
I've known him for a long time, this director is one of the best in his field.	Lo conosco da molto tempo, questo regista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo regista è uno dei migliori nel suo campo.
I've known her for a long time, this director is one of the best in her field.	La conosco da molto tempo, questo regista è uno dei migliori nel suo campo.	La conosco da molto tempo, questo regista è uno dei migliori nel suo campo.
I've known him for a long time, this tour guide is one of the best in his field.	Lo conosco da molto tempo, questa guida turistica è una delle migliori nel suo campo.	Lo conosco da molto tempo, questa guida turistica è una delle migliori nel suo campo.
I've known her for a long time, this tour guide is one of the best in her field.	La conosco da molto tempo, questa guida turistica è una delle migliori nel suo campo.	La conosco da molto tempo, questa guida turistica è una delle migliori nel suo campo.
I've known him for a long time, this bus driver is one of the best in his field.	Lo conosco da molto tempo, questo autista di autobus è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo autista di autobus è uno dei migliori nel suo campo.
I've known her for a long time, this bus driver is one of the best in her field.	La conosco da molto tempo, questo autista di autobus è uno dei migliori nel suo campo.	La conosco da molto tempo, questo autista di autobus è uno dei migliori nel suo campo.

I've known him for a long time, this professor is one of the best in his field.	Lo conosco da molto tempo, questo professore è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo professore è uno dei migliori nel suo campo.
I've known her for a long time, this professor is one of the best in her field.	La conosco da molto tempo, questo professore è uno dei migliori nel suo campo.	La conosco da molto tempo, questa professoressa è una delle migliori nel suo campo.
I've known him for a long time, this scientist is one of the best in his field.	Lo conosco da molto tempo, questo scienziato è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo scienziato è uno dei migliori nel suo campo.
I've known her for a long time, this scientist is one of the best in her field.	La conosco da molto tempo, questa scienziata è una delle migliori nel suo campo.	La conosco da molto tempo, questa scienziata è una delle migliori nel suo campo.
I've known him for a long time, this photographer is one of the best in his field.	Lo conosco da molto tempo, questo fotografo è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo fotografo è uno dei migliori nel suo campo.
I've known her for a long time, this photographer is one of the best in her field.	La conosco da molto tempo, questa fotografa è una delle migliori nel suo campo.	La conosco da molto tempo, questa fotografa è una delle migliori nel suo campo.

Input originale (subset femminile)	Output Deepl	Output Google Translate
I've known him for a long time, this judge is one of the best in his field.	Lo conosco da molto tempo, questo giudice è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo giudice è uno dei migliori nel suo campo.
I've known her for a long time, this judge is one of the best in her field.	La conosco da molto tempo, questo giudice è uno dei migliori nel suo campo.	La conosco da molto tempo, questo giudice è uno dei migliori nel suo campo.
I've known him for a long time, this magistrate is one of the best in his field.	Lo conosco da molto tempo, questo magistrato è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo magistrato è uno dei migliori nel suo campo.
I've known her for a long time, this magistrate is one of the best in her field.	La conosco da molto tempo, questo magistrato è uno dei migliori nel suo campo.	La conosco da molto tempo, questo magistrato è uno dei migliori nel suo campo.
I've known him for a long time, this bartender is one of the best in his field.	Lo conosco da molto tempo, questo barista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo barista è uno dei migliori nel suo campo.
I've known her for a long time, this bartender is one of the best in her field.	La conosco da molto tempo, questo barista è uno dei migliori nel suo campo.	La conosco da molto tempo, questa barista è una delle migliori nel suo campo.
I've known him for a long time, this artist is one of the best in his field.	Lo conosco da molto tempo, questo artista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo artista è uno dei migliori nel suo campo.

I've known her for a long time, this artist is one of the best in her field.	La conosco da molto tempo, questo artista è uno dei migliori nel suo campo.	La conosco da molto tempo, questa artista è una delle migliori nel suo campo.
I've known him for a long time, this writer is one of the best in his field.	Lo conosco da molto tempo, questo scrittore è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo scrittore è uno dei migliori nel suo campo.
I've known her for a long time, this writer is one of the best in her field.	La conosco da molto tempo, questa scrittrice è una delle migliori nel suo campo.	La conosco da molto tempo, questa scrittrice è una delle migliori nel suo campo.
I've known him for a long time, this baker is one of the best in his field.	Lo conosco da molto tempo, questo panettiere è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo fornaio è uno dei migliori nel suo campo.
I've known her for a long time, this baker is one of the best in her field.	La conosco da molto tempo, questo panettiere è uno dei migliori nel suo campo.	La conosco da molto tempo, questo fornaio è uno dei migliori nel suo campo.
I've known him for a long time, this pharmacist is one of the best in his field.	Lo conosco da molto tempo, questo farmacista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo farmacista è uno dei migliori nel suo campo.
I've known her for a long time, this pharmacist is one of the best in her field.	La conosco da molto tempo, questa farmacista è una delle migliori nel suo campo.	La conosco da molto tempo, questa farmacista è una delle migliori nel suo campo.
I've known him for a long time, this accountant is one of the best in his field.	Lo conosco da molto tempo, questo commercialista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo commercialista è uno dei migliori nel suo campo.

I've known her for a long time, this accountant is one of the best in her field.	La conosco da molto tempo, questa contabile è una delle migliori nel suo campo.	La conosco da molto tempo, questa contabile è una delle migliori nel suo campo.
I've known him for a long time, this appraiser is one of the best in his field.	Lo conosco da molto tempo, questo perito è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo perito è uno dei migliori nel suo campo.
I've known her for a long time, this appraiser is one of the best in her field.	La conosco da molto tempo, questo perito è uno dei migliori nel suo campo.	La conosco da molto tempo, questo perito è uno dei migliori nel suo campo.
I've known him for a long time, this veterinarian is one of the best in his field.	Lo conosco da molto tempo, questo veterinario è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo veterinario è uno dei migliori nel suo campo.
I've known her for a long time, this veterinarian is one of the best in her field.	La conosco da molto tempo, questo veterinario è uno dei migliori nel suo campo.	La conosco da molto tempo, questa veterinaria è una delle migliori nel suo campo.
I've known him for a long time, this tax collector is one of the best in his field .	Lo conosco da molto tempo, questo esattore delle tasse è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo esattore delle tasse è uno dei migliori nel suo campo.
I've known her for a long time, this tax collector is one of the best in her field.	La conosco da molto tempo, questo esattore delle tasse è uno dei migliori nel suo campo.	La conosco da molto tempo, questo esattore delle tasse è uno dei migliori nel suo campo.

I've known him for a long time, this health worker is one of the best in his field. I've known her for a long time, this health worker is one of the best in her field.	Lo conosco da molto tempo, questo operatore sanitario è uno dei migliori nel suo campo. La conosco da molto tempo, quest'operatore sanitario è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo operatore sanitario è uno dei migliori nel suo campo. La conosco da molto tempo, questo operatore sanitario è uno dei migliori nel suo campo.
I've known him for a long time, this cashier is one of the best in his field.	Lo conosco da molto tempo, questo cassiere è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo cassiere è uno dei migliori nel suo campo.
I've known her for a long time, this cashier is one of the best in her field.	La conosco da molto tempo, questa cassiera è una delle migliori nel suo campo.	La conosco da molto tempo, questa cassiera è una delle migliori nel suo campo.
I've known him for a long time, this model is one of the best in his field.	Lo conosco da molto tempo, questo modello è uno dei migliori nel suo campo.	Lo conosciamo da molto tempo, questo modello è uno dei migliori nel suo campo.
I've known her for a long time, this model is one of the best in her field.	La conosco da molto tempo, questa modella è una delle migliori nel suo campo.	La conosco da molto tempo, questo modello è uno dei migliori nel suo campo.
I've known him for a long time, this tailor is one of the best in his field.	Lo conosco da molto tempo, questo sarto è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo sarto è uno dei migliori nel suo campo.

I've known her for a long time, this tailor is one of the best in her field.	La conosco da molto tempo, questo sarto è uno dei migliori nel suo campo.	La conosco da molto tempo, questo sarto è uno dei migliori nel suo campo.
I've known him for a long time, this psychologist is one of the best in his field.	Lo conosco da molto tempo, questa psicologa è una delle migliori nel suo campo.	Lo conosco da molto tempo, questo psicologo è uno dei migliori nel suo campo.
I've known her for a long time, this psychologist is one of the best in her field.	La conosco da molto tempo, questa psicologa è una delle migliori nel suo campo.	La conosco da molto tempo, questa psicologa è una delle migliori nel suo campo.
I've known him for a long time, this librarian is one of the best in his field.	Lo conosco da molto tempo, questo bibliotecario è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo bibliotecario è uno dei migliori nel suo campo.
I've known her for a long time, this librarian is one of the best in her field.	La conosco da molto tempo, questa bibliotecaria è una delle migliori nel suo campo.	La conosco da molto tempo, questa bibliotecaria è una delle migliori nel suo campo.
I've known him for a long time, this teacher is one of the best in his field.	Lo conosco da molto tempo, questo insegnante è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo insegnante è uno dei migliori nel suo campo.
I've known her for a long time, this teacher is one of the best in her field.	La conosco da molto tempo, questa insegnante è una delle migliori nel suo campo.	La conosco da molto tempo, questa insegnante è una delle migliori nel suo campo.

I've known him for a long time, this flight attendant is one of the best in his field.	Lo conosco da molto tempo, questa assistente di volo è una delle migliori nel suo campo.	Lo conosco da molto tempo, questo assistente di volo è uno dei migliori nel suo campo.
I've known her for a long time, this flight attendant is one of the best in her field.	La conosco da molto tempo, questa assistente di volo è una delle migliori nel suo campo.	La conosco da molto tempo, questa assistente di volo è una delle migliori nel suo campo.
I've known him for a long time, this social worker is one of the best in his field.	Lo conosco da molto tempo, questo assistente sociale è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo assistente sociale è uno dei migliori nel suo campo.
I've known her for a long time, this social worker is one of the best in her field.	La conosco da molto tempo, questa assistente sociale è una delle migliori nel suo campo.	La conosco da molto tempo, questa assistente sociale è una delle migliori nel suo campo.
I've known him for a long time, this teller is one of the best in his field.	Lo conosco da molto tempo, questo cassiere è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo cassiere è uno dei migliori nel suo campo.
I've known her for a long time, this teller is one of the best in her field.	La conosco da molto tempo, questa cassiera è una delle migliori nel suo campo.	La conosco da molto tempo, questa cassiera è una delle migliori nel suo campo.
I've known him for a long time, this occupational therapist is one of the best in his field.	Lo conosco da molto tempo, questo terapeuta occupazionale è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo terapeuta occupazionale è uno dei migliori nel suo campo.

I've known her for a long time, this occupational therapist is one of the best in her field.	La conosco da molto tempo, questa terapeuta occupazionale è una delle migliori nel suo campo.	La conosco da molto tempo, questa terapeuta occupazionale è una delle migliori nel suo campo.
I've known him for a long time, this nurse is one of the best in his field.	Lo conosco da molto tempo, questa infermiera è una delle migliori nel suo campo.	Lo conosco da molto tempo, questa infermiera è una delle migliori nel suo campo.
I've known her for a long time, this nurse is one of the best in her field.	La conosco da molto tempo, questa infermiera è una delle migliori nel suo campo.	La conosco da molto tempo, questa infermiera è una delle migliori nel suo campo.
I've known him for a long time, this cleaner is one of the best in his field.	Lo conosco da molto tempo, questa donna delle pulizie è una delle migliori nel suo campo.	Lo conosco da molto tempo, questo detergente è uno dei migliori nel suo campo.
I've known her for a long time, this cleaner is one of the best in her field.	La conosco da molto tempo, questa donna delle pulizie è una delle migliori nel suo campo.	La conosco da molto tempo, questa donna delle pulizie è una delle migliori nel suo campo.
I've known him for a long time, this nutritionist is one of the best in his field.	Lo conosco da molto tempo, questo nutrizionista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo nutrizionista è uno dei migliori nel suo campo.
I've known her for a long time, this nutritionist is one of the best in her field.	La conosco da molto tempo, questa nutrizionista è una delle migliori nel suo campo.	La conosco da molto tempo, questa nutrizionista è una delle migliori nel suo campo.

I've known him for a long time, this hairdresser is one of the best in his field.	Lo conosco da molto tempo, questo parrucchiere è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo parrucchiere è uno dei migliori nel suo campo.
I've known her for a long time, this hairdresser is one of the best in her field.	La conosco da molto tempo, questa parrucchiera è una delle migliori nel suo campo.	La conosco da molto tempo, questa parrucchiera è una delle migliori nel suo campo.
I've known him for a long time, this speech therapist is one of the best in his field.	Lo conosco da molto tempo, questo logopedista è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo logopedista è uno dei migliori nel suo campo.
I've known her for a long time, this speech therapist is one of the best in her field.	La conosco da molto tempo, questa logopedista è una delle migliori nel suo campo.	La conosco da molto tempo, questa logopedista è una delle migliori nel suo campo.
I've known him for a long time, this dental hygienist is one of the best in his field.	Lo conosco da molto tempo, questo igienista dentale è uno dei migliori nel suo campo.	Lo conosco da molto tempo, questo igienista dentale è uno dei migliori nel suo campo.
I've known her for a long time, this dental hygienist is one of the best in her field.	La conosco da molto tempo, questa igienista dentale è una delle migliori nel suo campo.	La conosco da molto tempo, questa igienista dentale è uno dei migliori nel suo campo.